

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:06:57 ; Search time 5315.66 Seconds
(without alignments)
10117.650 Million cell updates/sec

Title: US-09-802-208B-1
Perfect score: 1848
Sequence: 1 atgacgaacaattacatg.....tctacggttaataactaa 1848

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1848	100.0	1848	6	AX268026	AX268026 Sequence
C 2	860.8	46.6	5930	1	AF045245	AF045245 Klebsiell
C 3	588.6	31.9	10799	1	AE013803	AE013803 Yersinia
C 4	588.6	31.9	313050	1	AJ414152	AJ414152 Yersinia
C 5	412.2	22.3	200050	1	AL646068	AL646068 Ralstonia
C 6	118.2	6.4	10031	1	AF007800	AF007800 Pseudomon
C 7	113.2	6.1	349116	1	AF003003	AP003003 Mesorhizo
C 8	113	6.1	3250	8	SCYNR073C	Z71688 S.cerevisia
C 9	113	6.1	4931	8	SCYNR074C	Z71689 S.cerevisia
C 10	113	6.1	9172	8	SCS288C14	X86790 S.cerevisia
C 11	112	6.1	11164	1	AE009374	AE009374 Agrobacte
C 12	112	6.1	14945	1	AE008240	AE008240 Agrobacte
C 13	111.4	6.0	50849	8	SCS288C14	AE009374 Agrobacte
C 14	108.4	5.9	10449	1	AE005449	U18795 Saccharomyc
C 15	108.4	5.9	26857	1	AP002560	U18795 Saccharomyc
C 16	105	5.7	4073	1	AE028937	AE028937 Gluconoba
C 17	104	5.6	24336	1	AE008841	AE008841 Salmonell
C 18	103.2	5.6	10446	1	AE000306	AE000306 Escherich
C 19	103.2	5.6	17093	1	D90849	D90849 E.coli geno
C 20	103.2	5.6	75888	1	ECOHU47	U00007 47 to 48 ce
C 21	103.2	5.6	244894	2	AC084324	AC084324 Mus muscu
C 22	103	5.6	25034	1	AE008844	AE008844 Salmonell
C 23	103	5.6	230050	1	AL627277	AL627277 Salmonell
C 24	102.2	5.5	14681	1	AE013893	AE013893 Yersinia
C 25	102.2	5.5	199050	1	AJ414147	AJ414147 Yersinia
C 26	101.8	5.5	1458	1	AB063188	AB063188 Gluconoba
C 27	101.8	5.5	1458	6	AX012443	AX012443 Sequence
C 28	101.8	5.5	1458	6	E50685	E50685 Gene manipu
C 29	101	5.5	189370	1	AF010496	AF010496 Rhodobact
C 30	97.6	5.3	11258	1	AE000251	AE000251 Escherich
C 31	97.6	5.3	14045	1	D90797	D90797 E.coli geno
C 32	97.6	5.3	19465	1	D90798	D90798 E.coli geno
C 33	96.8	5.2	14050	1	EC305143	AJ305143 Erwinia c
C 34	96	5.2	11783	1	AE005350	AE005350 Escherich
C 35	96	5.2	286485	1	AP002557	AP002557 Escherich
C 36	87.6	4.7	9810	1	AF018073	AF018073 Rhodobact
C 37	84.4	4.6	10955	1	AE005663	AE005663 Escherich
C 38	84.4	4.6	168764	1	AP002569	AP002569 Escherich
C 39	84	4.5	4539	1	D13329	D13329 Escherichia
C 40	84	4.5	11277	1	AE000503	AE000503 Escherich
C 41	84	4.5	338534	1	ECOUW93	U14003 Escherichia
C 42	83.4	4.5	10788	1	AE012076	AE012076 Xanthomon
C 43	82.6	4.5	10977	1	AE004660	AE004660 Pseudomon
C 44	79.6	4.3	21585	1	AE008765	AE008765 Salmonell
C 45	79.2	4.3	989	11	CNS06JD2	AL401436 T7 end of

ALIGNMENTS

RESULT 1
AX268026
LOCUS AX268026
DEFINITION Sequence 1 from Patent WO0166779.
ACCESSION AX268026
VERSION AX268026.1 GI:16516559
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Parrott,W., Lafayette,P. and Kane,P.
TITLE Arabitol or ribitol as positive selectable markers
JOURNAL Patent: WO 0166779-A 1 13-SEP-2001;

University of Georgia Research Foundation, Inc. (US)

FEATURES Location/Qualifiers

source 1..1848 /organism="Escherichia coli"

BASE COUNT 494 a 443 c 473 g 438 t

ORIGIN

Query Match 100.0%; Score 1848; DB 6; Length 1848;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1848; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAACAATTTACATGGCTGCACATCGGGTAGGTTCTTTTCATCGCGCACATCAG 60
DB 1 ATGACGAACAATTTACATGGCTGCACATCGGGTAGGTTCTTTTCATCGCGCACATCAG 60
QY 61 GCGTGGTATCTACACCGTTTTCAGAGTATGAGCGGATAAACCGCTGGAGCATTTGCTGGGGC 120
DB 61 GCGTGGTATCTACACCGTTTTCAGAGTATGAGCGGATAAACCGCTGGAGCATTTGCTGGGGC 120
QY 121 AATATTCGTAAATGATGCTGAACATGCTGTACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 180
DB 121 AATATTCGTAAATGATGCTGAACATGCTGTACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 180
QY 181 GTGCTGAAACCGTCAGCCCGGAAGGGTAAGCGAATATGAAGAGATCACTCAATTCAG 240
DB 181 GTGCTGAAACCGTCAGCCCGGAAGGGTAAGCGAATATGAAGAGATCACTCAATTCAG 240
QY 241 AAGTTGATACCGTGGCAGGAGATTTACAACCGCTGATTTGCTGAAGGGGAGATCCGAAG 300
DB 241 AAGTTGATACCGTGGCAGGAGATTTACAACCGCTGATTTGCTGAAGGGGAGATCCGAAG 300
QY 301 ACAAAAGTATGCTTTTACCGTCACCGAAGGGGTACTACCTGTAATACCAAGTCAAAA 360
DB 301 ACAAAAGTATGCTTTTACCGTCACCGAAGGGGTACTACCTGTAATACCAAGTCAAAA 360
QY 361 CTGGAAGTTAACATCTGATTTAGCGGCAGATCTTAAAGGGGATGCAAAATTTTAC 420
DB 361 CTGGAAGTTAACATCTGATTTAGCGGCAGATCTTAAAGGGGATGCAAAATTTTAC 420
QY 421 GGTGTTATACCGTATCTCGAAGCGGTATGCAAAATACCGCGGACCACTAACCCCTG 480
DB 421 GGTGTTATACCGTATCTCGAAGCGGTATGCAAAATACCGCGGACCACTAACCCCTG 480
QY 481 ATGACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG 540
DB 481 ATGACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG 540
QY 541 GCGTGGTATCTACACCGTTTTCAGAGTATGCGCGATAAACCGCTGGAGCATTTGCTGGGGC 600
DB 541 GCGTGGTATCTACACCGTTTTCAGAGTATGCGCGATAAACCGCTGGAGCATTTGCTGGGGC 600
QY 601 AATATTCGTAAATGATGCTGAACATGCTGTACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 660
DB 601 AATATTCGTAAATGATGCTGAACATGCTGTACAGGCACTCAGTGCACAGAAAGGTTCGCTAT 660
QY 661 GTGCTGAAACCGTCAGCCCGGAAGGGTAAGCGAATATGAAGAGATCACTCAATTCAG 720
DB 661 GTGCTGAAACCGTCAGCCCGGAAGGGTAAGCGAATATGAAGAGATCACTCAATTCAG 720
QY 721 AAGTTGATACCGTGGCAGGAGATTTACAACCGCTGATTTGCTGAAGGGGAGATCCGAAG 780
DB 721 AAGTTGATACCGTGGCAGGAGATTTACAACCGCTGATTTGCTGAAGGGGAGATCCGAAG 780
QY 781 ACAAAAGTATGCTTTTACCGTCACCGAAGGGGTACTACCTGAATACCACTCAAAA 840
DB 781 ACAAAAGTATGCTTTTACCGTCACCGAAGGGGTACTACCTGAATACCACTCAAAA 840
QY 841 CTGGAAGTTAACATCTGATTTAGCGGCAGATCTTAAAGGGGATGCAAAATTTTAC 900
DB 841 CTGGAAGTTAACATCTGATTTAGCGGCAGATCTTAAAGGGGATGCAAAATTTTAC 900
QY 901 GGTGTTATACCGTATCTCTGGAAGCGGTATGGCAATTAACCGGACCACTAACCCCTG 960

Db 901 GGTGTTATACCGTATCTCTGGAAGCGGTATGGCAAAATACCGGACCACTAACCCCTG 960
QY 961 CTGAATTTGCCATAACGTCGCCCATTAATGTAAGCTTTTCCATGATGCCCTGGTTGAGTTT 1020
Db 961 CTGAATTTGCCATAACGTCGCCCATTAATGTAAGCTTTTCCATGATGCCCTGGTTGAGTTT 1020
QY 1021 CTCAGCTAACTGGCAACAGGATGTCTAGCTGCTGCGGACGACAACTTCGGGACGGATCAAG 1080
Db 1021 CTCAGCTAACTGGCAACAGGATGTCTAGCTGCTGCGGACGACAACTTCGGGACGGATCAAG 1080
QY 1081 AATACCATGTTGACCGCATTTAGCCCTGCTGCGGACGACAACTTCGGGACGGATCAAG 1140
Db 1081 AATACCATGTTGACCGCATTTAGCCCTGCTGCGGACGACAACTTCGGGACGGATCAAG 1140
QY 1141 GCTCAACAGGGTATTTGCCGATAAAGCCCGGTAAATGGCGCAAACTTTTATCCAGTGGTC 1200
Db 1141 GCTCAACAGGGTATTTGCCGATAAAGCCCGGTAAATGGCGCAAACTTTTATCCAGTGGTC 1200
QY 1201 GTGGAAGATAATTTCCGCTGATGTCGCTGCGGACATGGAGAAGTTCGCTGCAACATGGT 1260
Db 1201 GTGGAAGATAATTTCCGCTGATGTCGCTGCGGACATGGAGAAGTTCGCTGCAACATGGT 1260
QY 1261 GCGTCGCTAATCCCTATGAAGAGGCGCAAGATTCGCAATTTTAACTCTTCCACAGTTGC 1320
Db 1261 GCGTCGCTAATCCCTATGAAGAGGCGCAAGATTCGCAATTTTAACTCTTCCACAGTTGC 1320
QY 1321 ATGCGCTGGCAGGTACGTTTAAATCGGTCAAAAATATATCCACGAAAGCAACAATGACCGAT 1380
Db 1321 ATGCGCTGGCAGGTACGTTTAAATCGGTCAAAAATATATCCACGAAAGCAACAATGACCGAT 1380
QY 1381 TTTATCTATCAGATTTGCCGACCGCTACGTGACGGAAGATGTCAATTCCTTGGGGGAT 1440
Db 1381 TTTATCTATCAGATTTGCCGACCGCTACGTGACGGAAGATGTCAATTCCTTGGGGGAT 1440
QY 1441 AAGCGTATCAGATTTGCCAACCCTACCGTGATTTGTACTCAAGCGTTTACCAGTCCACAT 1500
Db 1441 AAGCGTATCAGATTTGCCAACCCTACCGTGATTTGTACTCAAGCGTTTACCAGTCCACAT 1500
QY 1501 ATTTCAGGACACCAACCAACCGCTGCGGTGCGGATGTTTCTCGAAATTTCCGGCGATGAT 1560
Db 1501 ATTTCAGGACACCAACCAACCGCTGCGGTGCGGATGTTTCTCGAAATTTCCGGCGATGAT 1560
QY 1561 GCCCCACACTGCGAGAGTGTCTACAGGAGGCGTTGCGCCGATGCCACCGCATGTTA 1620
Db 1561 GCCCCACACTGCGAGAGTGTCTACAGGAGGCGTTGCGCCGATGCCACCGCATGTTA 1620
QY 1621 CCGTCACTGTTTACGTTATTCATGGAGCGAGTGGCATCACGGCAACCTGCGCTATGAATAT 1680
Db 1621 CCGTCACTGTTTACGTTATTCATGGAGCGAGTGGCATCACGGCAACCTGCGCTATGAATAT 1680
QY 1681 CAGGATGGCATCCTTGTATGCACGAGTGTCCATGCAATGTTTACAGTCTGCCGATCCCGTC 1740
Db 1681 CAGGATGGCATCCTTGTATGCACGAGTGTCCATGCAATGTTTACAGTCTGCCGATCCCGTC 1740
QY 1741 GCTGTTTATGCCAGTATAAAGCGCTGTTTGGCGATTTTAAACGAACTGAAGATTTTGC 1800
Db 1741 GCTGTTTATGCCAGTATAAAGCGCTGTTTGGCGATTTTAAACGAACTGAAGATTTTGC 1800
QY 1801 GCGTTGTTGGCGGAAATTCGCTGAGCTACCGTGTAAATTAACATAA 1848
Db 1801 GCGTTGTTGGCGGAAATTCGCTGAGCTACCGTGTAAATTAACATAA 1848

RESULT 2

AF045245/c

LOCUS

DEFINITION

AF045245

Klebsiella pneumoniae D-arabinitol transporter (dalT),

D-xylulose-kinase (dalk), D-arabinitol dehydrogenase (dalD), and

repressor (dalR) genes, complete cds.

AF045245 U97126

AF045245.1 GI:2905644

VERSION

KEYWORDS

5930 bp DNA linear BCT 30-AUG-2001

QY	1081	AATACCATGTTTACCGCATTTACCGCTCTGTCGGCAGCAGAACTTCGGGCACCGATCAAG	1140
Db	4009	AACACCATGTTGACCGCATACACCCCGCTGTCGGCGCCGATGTCGCCGCCATCAAA	3950
QY	1141	GCTCAACAGGGTATTGCGGATAAGCCGCGTAAATGGCGAAACCTTTATCCAGTGGGTC	1200
Db	3949	GCCGACCGGATTTGATGACAAAGCCGCGTATGGGGAGACCTTTATCCAGTGGGTA	3890
QY	1201	GTGGAAGATAATTTCCGTGATGTCGTCGGCAGCTGGAGAAGTGTGTCGAACATGGTG	1260
Db	3889	GTGGAGAACAATTTCCCGCATGTCGCCCGCAATCTCGAGGCACTCGCGGTGGAGATGGTG	3830
QY	1261	CGTCGGTAATCCCTATGAAGAGGAGAGATTCGATTCCTTAACCTTTCACACACTGGC	1320
Db	3829	GAGTCGGTCATCCCGTATGAAGAGGGAATCCGATTTTGAACGCTCGCACAGCTGC	3770
QY	1321	ATCCCTGGCAGGTACGTTAATCGTCAAAAATATATCCACGAAAGCAATGACCGCAT	1380
Db	3769	ATTGCTTGGCGGAACCTTAATCGCCAGCAGTATATTATGAAGACGCTGACCGCAT	3710
QY	1381	TTTATCATCATGATTCCGACCCGCTACGTGACGGAAGATGTCTATTCCTTGTGGCGCAT	1440
Db	3709	GTTATCTACGCCATTCGCCAGCCGCTACGTTACGAGGACGTTATTCCGTCTCTCGGCGAC	3650
QY	1441	AACGGTATCGATTTGCCAACCTACCGTGATGTTGTACTCAAGCGTTTACCATCCACAT	1500
Db	3649	AACGGCATTTGATCTCGCGACCTACCGGGATGTGTGCTTAAAGCGCTTTACCAACCCCTAT	3590
QY	1501	ATTTCAGGACACCAACACGCGTCGTCGGGATGTTTCTCGAAAATTCGGCGGATGATT	1560
Db	3589	ATCCAGGACACCAACGCGCTGCGCGCGGCGGCTTCTCGAAAATTCGGCGGATGATC	3530
QY	1561	GCCCCACACATCGGAGAGTGCTACCAAGCAGGCGTTTCGCCGAATGCCACCGCATGTTA	1620
Db	3529	GCGCCAACCTTCGAGAGTGCTACCAAGCAGGCGTCCGCCGGAAGCGCGCATGCTG	3470
QY	1621	CCTGCACTGTTTACGTATTCATGAGCAGTAGTGATCAGGCAAACTGCCATGATGATAT	1680
Db	3469	CGGCGCTGTTCTCTCTTTATGAGCAGTAGTGATCAGGCAAACTGCCATATAGTATC	3410
QY	1681	CAGGATGGCATCTTGATGCACCACTGTCATGCAATGTTACAGTCTGCCGATCCCGTC	1740
Db	3409	CAGGACGGCATCTGATGCGGAGCGGTGCATGAGATGTTTCGAGGCTCAGGATCCGGTG	3350
QY	1741	CTGTTTATGCCAGTGAATAAGCGCTGTTTGGCGATTAAACGAAAGTGAAGATTTGCC	1800
Db	3349	GCTGTCTGCGCGTGATAAAGCCCTGTTTGGTGATCTGCCAACACGCGGATTTCCGTG	3290
QY	1801	GCCTGCTTGGCGGAAATAATCGCTGACGTCTACGCGCTTAATTAACATAA	1848
Db	3289	GCCTGATGCGGAGAAAGTCGCCCGCGTTTACACGCTGATTAACTAA	3242
RESULT 3			
AE013803/c			
LOCUS	AE013803	10799 bp	DNA linear BCT 26-JUL-2002
DEFINITION	Yersinia pestis KIM section 203 of 415 of the complete genome.		
ACCESSION	AE013803	AE009952	
VERSION	AE013803.1	GI:21958827	
KEYWORDS			
SOURCE	Yersinia pestis KIM.		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.		
REFERENCE	1 (bases 1 to 10799)		
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.		
TITLE	Genome Sequence of Yersinia pestis KIM		
JOURNAL	J. Bacteriol. 184 (16), 4601-4611 (2002)		
PUBMED	12142430		

REFERENCE	2 (bases 1 to 10799)		
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		
FEATURES	Location/Qualifiers		
source	1..10799		
	/organism="Yersinia pestis KIM"		
	/strain="KIM"		
gene	/db_xref="taxon:187410"		
	267..605		
	/gene="emrE"		
	/note="y2000"		
CDS	267..605		
	/gene="emrE"		
	/function="transport; drug/analog sensitivity"		
	/note="residues 3 to 112 of 112 are 50.90 pct identical to residues 1 to 110 of 110 from E. coli K12 : B0543;		
	residues 3 to 112 of 112 are 54.54 pct identical to residues 1 to 110 of 110 from GenPept : >gb AAC64463.1		
	(U67194) QacE [Enterobacter aerogenes]"		
	/codon_start=1		
	/transl_table=11		
	/product="methylviologen resistance"		
	/protein_id="AA085566.1"		
	/db_xref="GI:21958828"		
	/translation="MIMSFAVYLFMAITAEVATTMLKASDGFSLRVPVVVIGYI		
	AFWGLSQVVKTMPLGIAYIAVSLGLVLSIAATTFMYQKLDMAAVVGNALIIISVMV		
	INLLSKTPMH"		
	complement(665..1324)		
gene	/gene="y2001"		
	/complement(665..1324)		
CDS	/gene="y2001"		
	/function="putative transport"		
	/note="residues 1 to 195 of 219 are 50.25 pct identical to residues 1 to 195 of 215 from E. coli K12 : B3508"		
	/codon_start=1		
	/transl_table=11		
	/product="putative transport ATPase"		
	/protein_id="AA085567.1"		
	/db_xref="GI:21958829"		
	/translation="MITDLRLITLAGLIGLQLRKAEGILRTHILVIGSAM		
	FMIVSKYGFEDLLVLEHVSFDPSSVAQVWVGMLGACTIMIQKVMKGLTAAAGMW		
	VTAAIGLVIGSLYEIGIYGLTMTLVLEVFRLQSRLGLGHQYLVVLLPSPVLVL		
	VALQKLHLPMLSVTONHKSDNTSELTMEVTLSPKTVNQVYEQLLAIEGHIQLDI		
	K"		
	1537..1950		
gene	/gene="y2002"		
	1537..1950		
CDS	/gene="y2002"		
	/function="unknown"		
	/note="residues 87 to 134 of 137 are 56.25 pct identical to residues 1 to 48 of 51 from GenPept : >dbj BAAL4988.1		
	(D90776) ORF_ID:02646; similar to [SwissProt Accession Number P52646] [Escherichia coli]"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical"		
	/protein_id="AA085568.1"		
	/db_xref="GI:21958830"		
	/translation="NMVYIILAFINRYLRISLSCRKMRIFQGVVCGMALFLAACSS		
	NDNSTLNNSNTLANSYTEPEQCATNAIAYRVNVMNSNDVNMNSAANCANAGGIL		
	ANTRQLHGSVGMCPDQKRCDEALLRGACPVK"		
	complement(2008..2460)		
gene	/gene="hslJ"		
	/note="y2003"		
	complement(2008..2460)		
CDS	/gene="hslJ"		
	/function="factor; adaptations, atypical conditions"		


```
QY 979 CGGCATAATAGGTGAACGTTTCCATGATGCGCTGGTTCAGTTTCAGCTTCCAGCTAACTGGCAAA 1038
Db 8153 CGTCACAACCGTGAACGATTCGCCCATGGCTTATTAGAGTTTCTTGCTCTCAGAGGTCAA 8094
QY 1039 CAGGATGTCTCAGCTGGCTGAGTACAAATACCACTTGCCTCGAATACCATGGTTGACCGC 1098
Db 8093 CAAGCGCTATTAACTGGGTGACAGCGACCGCGGCGAATACCATGGTTGATGCGT 8034
QY 1099 ATTACGCTCTCCGCGCAGACAGAACTTCGCGCACGAGTCAACGCTCAAAACCGGATATTGCC 1158
Db 8033 ATCACCCCGCGCCCATCAGCGGATATCGCGCAGCGGGTGTGGAATAACCGGTATCGTA 7974
QY 1159 GATAAGCGCGCGGTATATGGCGAAACCTTTATCCAGTGGGTCTGGAAGATAATTTCCCGT 1218
Db 7973 GATAAGCGCTCGGTATGGGGAATCATTTATTCAATGGGTTATTGAGGATGATTTATT 7914
QY 1219 GATGTCGCTCGCGCACTGGAGAGGTGCGGTGTCGAACCTGGTGGCGTGGTAAATCCCTAT 1278
Db 7913 GCTGGCGCGCGCACTGGAGAACGTTGGGGCGGAATGGTCACTCGGTTTGGCCCTAT 7854
QY 1279 GAAGAGCGGAAGATTCGCATTCTTCACTTTCACACAGTTTGCATCGCTGGCGAGGTACG 1338
Db 7853 GAAGAAGCCAAATCCGATTCTGAACGCCACGACATGTTGTTGCTGGCGCGGACG 7794
QY 1339 TTAATCGGTCAAAATATATCCAGGAAGACCAATGACCGATTTTATCTATCAGATTGCC 1398
Db 7793 TTGATTGGGCAATGTTATATTATCATGAAGTACCAACACGACGCTATCCGCCAGATGGCT 7734
QY 1399 GACCGGTACTGACGGAAGATGTCATTCTTCTGGCGGATACGATATCGATTGCGCA 1458
Db 7733 TATGACTATGTCACGAAGATGTCATCCCTCCTGAGCGCCAGCCACTTGATTAGCC 7674
QY 1459 ACCTACCGGTGTTGTACTCAAGCGTTTTTACCAATCCACATATTCAGGACACCAACCAA 1518
Db 7673 CATTAACCGGATGTTGTTACTGGAACGTTTCAGTAATCCGTTATTCGCGATACGAACCA 7614
QY 1519 CGCGTGGCTGGGATGTTCTCGAAATTCGCGGATGATTCGCCCGCACACATGCGAGAG 1578
Db 7613 CGTGGCGCGCGGATGATTTTCCAAATCCCGGATTTATCATCTCCGCGCTGATCGAG 7554
QY 1579 TGCTACAGCGAGCGGCTTCGCCGAATGCCCGCATGTTTACCTGCACTGTTTTCAGTA 1638
Db 7553 TGTATACCAAGTGGCGAGTGCTTACCGCTACCGCAATGTTGCCGCAATTATTTCTCGTA 7494
QY 1639 TTCAATGAGCAGTGGCATCAGCGCAACGCTGCTTCAATATCCAGGATGCAATCCTTGTAT 1698
Db 7493 TTTATGGAAGTTTGGCAACCAAGCACTCTGCCGTATGATATCAATATCAGGATGATTCGGAC 7434
QY 1699 GCACGAGCTGCTCATGCAATGTTTACAGTCTGCCGATCCCGTCTGTTTATGCCAGTGAT 1758
Db 7433 CCAGCAGCGCTGCTCATGCCATGTTTCACTCTCTGACCCCTGCAGGCTTATTTGCCCGGTAT 7374
QY 1759 AAAGCGCTGTTTGGCGATTTTAAACCGAAGTGAAGATTTTGCCTCCGTTGTTGCCGCAAAA 1818
Db 7373 AGCGCCCTTTTTCGGCGCGTGGCCATGATGATGATCAGTTCGAACAGTATTATACCCAGTCC 7314
QY 1819 ATCGCTGACCTTACCGGTTAAATTAAC 1845
Db 7313 ATAAATCGTCTCTCTGTTGGGTTAGC 7287

RESULT 4
AJ414152 313050 bp DNA linear BCT 06-JUN-2002
LOCUS Yersinia pestis strain C092 complete genome; segment 12/20.
DEFINITION AJ414152 AL590842
ACCESSION AJ414152.1 GI:15980308
VERSION
KEYWORDS Yersinia pestis.
SOURCE Yersinia pestis
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
REFERENCE 1 (bases 1 to 313050)
```

AUTHORS

Parikh, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B.G. Genome sequence of *Yersinia pestis*, the causative agent of plague Nature 413 (6855), 523-527 (2001)

TITLE

Genome sequence of *Yersinia pestis*, the causative agent of plague

MEDLINE

21470413

PUBMED

11586360

REFERENCE

2 (bases 1 to 313050)

AUTHORS

Parikh, J.

TITLE

Submitted on behalf of the Versinia

JOURNAL

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENT

Notes:

Details of *Y. pestis* sequencing at the Sanger Centre are available

on the World Wide Web.

(URL, http://www.sanger.ac.uk/projects/Y_pestis/).

FEATURES

Location/Qualifiers

1..313050

/organism="Yersinia pestis"

/strain="C092"

/db_xref="taxon:632"

/note="biovar: Orientalis"

21470413

/gene="YPO2312"

633..638

/gene="YPO2312"

645..3269

/gene="YPO2312"

/note="Similar to Photobacterium luminescens insecticidal

toxin complex protein Tcc TR:O85157 (EMBL:AF047028) (1043

aa) fasta scores: E(): 0.604% id in 695 aa, and to

Serratia entomophila plasmid PADAP virulence determinant

SePC TR:AG09644 (EMBL:AF135182) (973 aa) fasta scores:

E(): 0.622% id in 875 aa. Similar to YPO2380 (984 aa)

fasta scores: E(): 0.649% identity in 880 aa overlap,

YPO3674 (1011 aa) fasta scores: E(): 0.501% identity in

773 aa overlap, and to YPO3673 (952 aa) fasta scores: E():

0.521% identity in 729 aa overlap. Note the differing

C-terminus of the product of this CDS and the G+C content

-56%

/codon_start=1

/transl_table=11

/product="putative insecticidal toxin complex"

/protein_id="CAC91117.1"

/db_xref="GI:15980309"

/db_xref="SPTREMBL:Q8ZE71"

/translation="MRTSLFSTPSVTVLNRLGVTRVDIAYHRHDPDPVIERITHH

QYDARGFLTQSDAPRLHAGLMNFVSLDTLGRILTOGADNGTTVSLNDAAGRPIS

ASNIYSTSDGTEDRGQAMRTWQYBEASLPGLVSVTEQVTKATRIERFVVAANTD

AKSLNLGACVSHYDVLGVDPSLTGLVPLSVTRRLKMSADNPDAVADMGADAS

ANNOLDGETHSLRTADATGAVLTDDAKGNLQRMAYDAGLLSGSLTLKDGTEQV

IVKSLTSAAGKLEEHNGVVTIEYEPETQRLVGIKTERPAGHASAKVLQDLURY

EYDPVGNVLKVTNDAEETFRWQNVPPENTYDLSYOLVSATGEMANAGQSCSL

PSTYPLPADSSAYTRYSTYTDGANTQIRHNAPNTNSYTKITVSDRSNRGLV

STLTENAADVALFTAGGQTKLPQCHLWTRNELLKVTVPVRDGTSDSESRYD

AAORILKVSROKTNMTSMOTQRLVPLGLLELRKSKGDTETEGLOVITVGEAGQVR

VLHESGSPDEITDDOIRYSYDNLGACSLGELGDGNIISAEEYYPGCTAYWAVRA

VEADYKTVRSKERTADGLYYGYRYOPWAGRWLSADPWAGVGLNLFKRNPNPV

AFIDRNLNSSELLYQAFKATANKYVIGVAPNPPLGTLLEKGFPPSKHMKAKSS

PTGPTAGTAEADPTYSKSPSAYKKORASIDKALGSESIDLFTSKSRINELIDTGN

LNSLGNRYSAKYPTCTQTEFGNNGRVLNSEGKPKVPMNTNPPEIGERKSSNSPITAD

YDLFALIOVLISQLMKDP"

3272..3628

/gene="YPO2313"

3272..3628

/gene="YPO2313"

/note="No significant database hits"

/codon_start=1

/transl_table=11

QY	859	GATTTAGCGCAGATCTTAAAGGGGATGCAAAACAATTTACGGTGTATTATCCCGTATC	918
Db	15310	GATATTACGGCCGATTTAGAGAGATTCGCCACGATTTACGGGCAATTACCGCTATC	15369
QY	919	CTCGAAGCGGTATGCAAAATACCGCGACCACTAACCTGCTGAATTCGGAATAACGGT	978
Db	15370	TTCCACAACGGCAACAGGATAGAGTGGCCCGTCACTCTACTGAACTCGGAATACTTG	15429
QY	979	CGCCATAATGGTGAAGCTTTCCATGATGGCTGGTTGAGTTTCTCCAGCTAACTGCCAA	1038
Db	15430	CGTCACAACGGTCAACGATTTCCGCCATGGCTATTAGAGTTTCTGCTACGAGTCAA	15489
QY	1039	CAGGATGTATCAGCTGGCTGAGTACAAATACACATTTGCCGAATACCATGGTTGACCGC	1098
Db	15490	CAAAGCCTATTAAACTTGGGTGACAAGCCAGACCCGAGCCGAATACCATGGTTGATCGT	15549
QY	1099	ATTACGCTCGTCCGCGACGACAGAACTTCGCGCACGATCAACGCTCAAAAGGGTATTGCC	1158
Db	15550	ATCACCCGCGCCCATCAGCCGATATCGCCACGCGGTGTGGAANAACCGGTATCGTA	15609
QY	1159	GATAAAGCGCCGTAATGGCGCAACCTTTATCCAGTGGGTCTGGAAGATAATTTCCCGT	1218
Db	15610	GATAAGCCTCCGCTGATGGGGAATCATTTATTCAATGGGTATTGAGGATGATTTTATT	15669
QY	1219	GATGTCGTCGCGCACTGGAGAGTCCGTGTCGAACCTGGTGGCGTCGGTAATCCCGTAT	1278
Db	15670	GCTGGCGCCCGCGATCGGAGACGTTGGGGCGGAATGGTCACTCGGTTTGGCCTTAT	15729
QY	1279	GAAGAGCGGAAGATTCCGATTCCTTAACCTCTTACACAGTTGCACTCGCTGGCGAGTACG	1338
Db	15730	GAAGAAGCAAAATCCGATCTTGAACGCCAGCCACAGTTGATTGCTTGGCGGCGCAGC	15789
QY	1339	TTAATCGGTCAAAATATATCCAGAAAGACAAATGACCGGATTTATCTATCAGATGGCC	1398
Db	15790	TTGATTGGGCAATGTATATTTCATGAAGTACCAACCGCACGCTATCCGCCAGATGGCT	15849
QY	1399	GACCGCTACGTGACGAAGATGTCATTCTGCTTGGCGGATAACGGTATCGATTTGCCA	1458
Db	15850	TATGACTATGTACGCAAGATGTCATTCTCCCTCCCTGACGCCACGACCTTGATTTAGCC	15909
QY	1459	ACCTACCGTGATGTTGCTACTCAAGCGTTTTTACCAATCCACATATTCAGGACACCAACCA	1518
Db	15910	CATTACCGCGATGTTGCTACTGGAACGTTTCAGTAACTCGTATATTCGCGATACGAACCG	15969
QY	1519	CGCGTCGTCGCGATGGTTCTCGAAATTCGCGGATGATTCGCCCGACACATGCGAGAG	1578
Db	15970	CGTGTCCGCCCGATGGATTTTCCAAATCCCGGATTTATCACTCCACGCTGATCGAG	16029
QY	1579	TGCTACAGCGGCGGTTCGCCGAATGCCCGCATGTTACCTGCACCTGTTTTCAGTA	1638
Db	16030	TGTTACCAAGTGGCGAGTGCCCTACCGCTACCGCAATGTTGCCCGCATTTCTTCGTA	16089
QY	1639	TTCATGAGCAGTGGCATCACGCAACCTGCCCTATGAATATCAGGATGCATCTTGAT	1698
Db	16090	TTTATGAAAGTTGGCAACCAAGCACTCTGCCGTATGATATCAGGATGATTTCTGGAC	16149
QY	1699	GCACAGCTGTCCATGCAATGTATACGTTCGCGATTCGCCGCTGTTATGCGCACTGAT	1758
Db	16150	CCACAGCGCTCCATGCCATGTTTCACTCTCTGACCCCTGCAGGCTATTTTGCCCGTGT	16209
QY	1759	AAAGCGCTGTTTGGCGATTTACCGCAACGTGAAGATTTTGGCCCGTGTTCGCCGAAAA	1818
Db	16210	AGCGCCCTTTTGGGCGCGTGGCCAATGATGATCAGTTCGAACAGTTATTACGCCAGTCC	16269
QY	1819	ATCGCTGAGCTCTACCGGTTAATTAAC	1845
Db	16270	ATAAATCGTCTCTCTCTGGGTAGC	16296

RESULT 5
AL646068/c 200050 bp DNA linear BCT 07-DEC-2001
LOCUS Ralstonia solanacearum GM1000 chromosome, complete sequence;
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

COMMENT

FEATURES
source

gene

CDS

gene

CDS

segment 12/19.
AL646068 AL646052
AL646068.1 GI:17429060

Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.

1 (bases 1 to 200050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Sigulier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum

Unpublished
2 (bases 1 to 200050)
Boucher,C.A.
Direct Submission

Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CERH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Garbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

Location/Qualifiers.

1. .200050
/organism="Ralstonia solanacearum"
/strain="GM1000"
/db_xref="taxon:305"
complement(85..573)
/gene="RSc2042"
/note="RS03607"
complement(85..573)
/gene="RSc2042"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
protein_id="CAD15749.1"
/db_xref="GI:17429061"
/translation="MRVIESLDALRGLIGQEVAVSDWVEITQQOVNQFEATGDRDWI
HYDVERARRESFPGAPVANGFLTSLPALMQSALDPMVKMGVNTGLNKRFPAPVP
YSLRLRARSLSLEMEMLPLQSPDAPALPTGAQMTGWVTECEGQARPCVVAESISR
Y5"
complement(600..1070)
/gene="RSc2043"
/note="RS03608"
complement(600..1070)
/gene="RSc2043"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"

```
/protein_id="CAD15750.1"
/db_xref="GI:17429062"
/translation="MTEAALPYTFYFDFOVGOTMEMGTVAVTEDEILAFARQYDPOP
FHDPEARSISVIGLLSSGMMTCVAMRMVONFLSKSSSMGSPGVDEIRLWPVYP
GDILSVSTCLEVPPSSQSKPDRGVAINRWARNQHGLVCTLVGMGLFGRPPAA"
complement(1160..2290)
/gene="RSC2044"
/note="RSC3609"
complement(1160..2290)
/gene="RSC2044"
/note="RSC3609"
/EC_number="1.3.99.3"
/function="small molecule metabolism; degradation; fatty
acid"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE
PROTEIN"
/protein_id="CAD15751.1"
/db_xref="GI:17429063"
/translation="MDFSTDEOKQLADAVRRFIDKDYGFARNKVYVSPAGVSOAHW
DALAELGLTALPVPOAGGFDGSAMDLLVMQELGRGLVVEPYAATVLGMOALKAGS
QEALEVPAGALKLAARFGEPSRYELFNWTTAARQGGGWTLSGAKAVVHGGQAD
KLVSARTGGEARTSGLSLFDREAGVTVKDYRTIDNLRAADVDFNAPATLLGG
AGEWEITIDAVADIGCVLLCAEAIGLDALNAATLEYTKTRQQFVPIARFQALQHRM
VDMFIHAQOERSTYLAARHFDEGDAERARYVSAARVQQAAREVGEAVQLHGGM
GVTNELPAAMHFKRLTMTINTLGDVHHLARFASRPGFRDAA"
complement(2315..3511)
/gene="RSC2045"
/note="RSC3610"
complement(2315..3511)
/gene="RSC2045"
/EC_number="1.3.99.3"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE ACYL-COA DEHYDROGENASE OXIDOREDUCTASE
PROTEIN"
/protein_id="CAD15752.1"
/db_xref="GI:17429064"
/translation="MDLNVSAADDAFRQVRGWLAEELPKDIQDVNLNHRDLNRDDFV
RWHKALAAKSHSVPHVWCGTGWISQIKHIWDEECARIGAPGVLPFGVSNVAPVIM
RYGNEAKRYLPLRLDTCWMCQYSEPGSGDLASLKTAVREGDHYIVNGOKTWT
TLGQADMIICLVLPDQAKMOEGISFLIDMTKPGITVRPIMLDEHEVNEVFFDN
VRPAENLIGENRWGYAKYLGHGRTIARVGHRSKRELAFLKRALQHQNGRPLL
EDPVFAKVASLELMALETTVLRVVSSEGVGRPGPEASLLKIKGTQIQOMLTELM
VEAVGPQAPDPDYLEGETPRVATGNDAAFLAPYFENYKTSIYGGSNELQIRNIS
QMILGV"
complement(3592..4053)
/gene="RSC2046"
/note="RSC3611"
complement(3592..4053)
/gene="RSC2046"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
```

```
/protein_id="CAD15753.1"
/db_xref="GI:17429065"
/translation="MKVLDLRKANDHRFEWFGSDADLQSORGLLLTQPCVCGHTEVE
RLPSAPRLNLSSTDAAROSGTVKTSADDEGAPATLQORYLKAVRILLANTEDVGR
FAEAREMHYNAPERGIRGTTSHEEALELAEEGLEVNPLLVDPALKQPVH"
complement(4213..4830)
/gene="RSC2047"
/note="RSC3612"
complement(4213..4830)
/gene="RSC2047"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD15754.1"
/db_xref="GI:17429066"
/translation="MKFDPARLSADSNEAVSDPDAGLRRETOLASALVHQKFLTLKQD
VVRLPDGRNASREYLIHPGAVMMIPLFDDGVLMEROFYRYPVGKVMIEFFPAGKLDPRE
GALACGRLEREETGYVAQRWDFLTRIHPVISYSTEFIDLYLARDLQPGESALDEGEF
LETFIAPAGOLIDWVRGTRISDVKTIIIGVFWLEKILSGTWTPGDA"
complement(4873..5181)
/gene="RSC2048"
/note="RSC3613"
complement(4873..5181)
/gene="RSC2048"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD15755.1"
/db_xref="GI:17429067"
/translation="MGTSTGLFVIVLALICANLPFNQRYLALLPARWPCKPFWLRG
VELMAYAVVGLVGVGIESGQGNFASQGWQFYATACLMLVFAFPFTWQYLVRRQRH"
complement(5198..6670)
/gene="nuon"
/note="RSC2049; RSC3614"
complement(5198..6670)
/gene="nuon"
/EC_number="1.6.5.3"
Query Match 22.3%; Score 412.2; DB 1; Length 200050;
Best Local Similarity 57.3%; Pred. No. 2.3e-117;
Matches 769; Conservative 0; Mismatches 568; Indels 6; Gaps 1;
QY 501 GCTGCACATCGGGTAGGTTCTTTTCATCGGCACATCAGCGTGTATCTACACCGTTT 560
||||| ||| || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 94513 GCTGCATATTGGCGCGGATCTTCCATCGTCGCCACAGCGGTGATCTCTGCATCGCGT 94454
QY 561 GCAGGTGATGGG-----CGATAAACGCTGGAGCATTTGTCGGGGCAATATTCTGTAATGA 614
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94453 CAACCGCGCGTGCCTCCGCGGAGCGTGGAGCTGACGTCGGGACATCTCGCGACCA 94394
QY 615 TGCTGAACATGCTGACAGGCACTCAGTGCACAGAAAGTTCGTATGTCTGGAACCGT 674
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94393 CATGCACGCGACACTCGCGCGCTGGCCGCGCAGCAGGGTGTCTACACGCTCGAGACCGT 94334
QY 675 CAGCCCGGAAGGGTAAAGCAATATCAAGAGATCACCTCAATTCAGAACCGTG 734
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94333 CACGCCGCAAGCGGCGTGCATGAGACGATCCGCTCGATCGCGCGGGTGTGCCCTG 94274
```

```
QY 735 GCAGGAGATTTACACCGCTGATTCCTGAAGGGGAGATCCGACACAAAAGTATGTC 794
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94273 GTCCGCGATCTTGGCGCGCTCATCAATACGGGCGGATCCGCGTGCAGTCTGTC 94214

QY 795 TTTACCGTCACGAGGCGGTACTACCTGAATACCACTACAACTGGAAGTTAACAA 854
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94213 GTTTACGGTCACGAAGGCGGTATTACCTCGACGACGATCGTCTCGATGTCACGA 94154

QY 855 TCTGTATTAGGGGAGATCTTAAAGGGGATGCAAAATTTACGGTGTATTACCGC 914
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94153 TCCGATCTCGGCGGACCTCGGGGCGCCCTCGACGCTATACGGCGCATTCGCCCG 94094

QY 915 TATCTCGAAGCGCGTATGGCAAAATTAACCGCGGACACATAACCTCTGTAATTCGCA 974
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94093 GTTGCTCGCGAGCGCGGACGCGGAGCGCGGCGCTTCACGCTCAGAGTTGCGACAA 94034

QY 975 CGTGCCCATATAGTGTAACGTTTCCATGATGCGCTGTGAGTTCTCCAGCTACTCG 1034
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94033 TCTGCGCAGACGCGCGGCTTTTCGCGGGGATGCGCGGCTTCTCGCTGCGCGG 93974

QY 1035 CAAACAGGATGTATCGACTGGCTGAGTACAAATACCACTTCCCGAATACCATGTTGA 1094
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93973 CGACGAGCGCTCTCGCTGGTTCGATGCCAACGTATCGTGCCGAGCGGATGTCGA 93914

QY 1095 CCGCATACCGCTCTCGCGCAGCAGAACTTCGGGACGAGTCAAGCTCAACGCGTAT 1154
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93913 CGGATACGCGCGCGCGCAGCAGCAGCGCTGCGCAGCGCGTGCACGCGGCCACCGCGT 93854

QY 1155 TGCAGTAAAGCGCGCGTAAATGGGCAAACTTTATTCAGTGGTGTGGAAGATAATTT 1214
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93853 CGACGACCGCTCGCGGTCATGGCGGAATCTTTCATCCAAATGGTGATCGAGACAAT 93794

QY 1215 CCGTGATGTCGCTCGCGCAGTGGAGAGTTCGTTGCGAATGGTGGCTCGGTAATCCC 1274
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93793 CATCGCGCGCGCGCGATGGAGATCGCGGCGGAGATCGTCCGCGATGTGATCC 93734

QY 1275 CTATGAAGAGGAGATTCGATCTTAACTTCTCACAGTGTGATCGCTGGCGCAGG 1334
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93733 GTACGAAGAGGAGATCCGATCTGTAATCGACGACAGTGCATCGCTGGCGCGG 93674

QY 1335 TAGCTTAATCGGTCAAAAATATATCCAGAAAGCAATGACCGATTTATCTATCAGAT 1394
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93673 CAGCTCGCGGAGTCACTACATTCACGAAGGATGCGTGATCGCGCATTTATCGCT 93614

QY 1395 TCCGACCGCTACGTGACGGAAGATGTCATCTTCTGTTGGCGATTAACGGTATCGATTT 1454
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93613 CGCCTACGACTACGTACCGACGACGTGATTCCTCGCTGACGCGGCGCGTGGATCT 93554

QY 1455 GCCAACCTACCGTATGTTGTAAGCGTTTACCAATCCACATATTCAGGACACCAA 1514
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93553 CGAGCGCTATCGGGAGTCTGCTCGAGCGCTTCGGCAATCGGTAGTCTCGACACCAA 93494

QY 1515 CCAACCGCTCGCTGCGGATGTTTCTCGAAATTCGCGGATGATTGCCCGCACACTGG 1574
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93493 CCAGCGTTCGCGCGGATGCTTCTCGAAATTCGCCGCTTCATCGCACCGACGCTCG 93434

QY 1575 AGATGTCTACCGAGGCGTTCGCCGAATGCCACCGCATGTTACCTGCATGTTTAT 1634
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93433 CGAATGCTTTCGCGGCGCGCGATCCCGTTCGCGACCGCGGTGTTGCGCGCTGTTCT 93374

QY 1635 CGTATTCATGGACAGTGCATCAGGCAAACTGCCCTATGAATATCAGGATGGCATCT 1694
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93373 GGGCTTCCTTGAAGTGTGGCGCGCGGCACTTGCCTACGTATACACGAGCGGTGAT 93314

QY 1695 TGATGCACCACTGTCCATGCAATGTTACAGTCTGCCGATCCGCTGCTTTATGCCAG 1754
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93313 GGATGCGCGCGCGACGAGCATGTCGAGGCTCGGATTCGTTGCGCGCTTCTGTTTC 93254

QY 1755 TGATAAGCGCTGTTTGGCGATTTAACCGAAGCTGAAGATTTTGGCGCTGTTGCGCA 1814
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93253 AGATGCCCAATTTGGGGTTCGCTTGGCGCGCGACGCGCTTGTTCAGGCGGTGCTGC 93194

QY 1815 AAAATCGCTGAGCTACCGCT 1837
```

```
Db 93193 GGGACCGCGCGCGTGAAGCAT 93171

RESULT 6
AF007800
LOCUS
DEFINITION
Pseudomonas fluorescens mannitol operon, MtlE (mtlE), MtlF (mtlF),
MtlG (mtlG), MtlK (mtlK), mannitol dehydrogenase (mtlD), xylulose
kinase (mtlY) and fructokinase (mtlZ) genes, complete cds.
AF007800 139468
AF007800.1 GI:2293413
Pseudomonas fluorescens.
Pseudomonas fluorescens
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 10031)
Brunker,P., Altenbuchner,J., Kulbe,K.D. and Mattes,R.
Cloning, nucleotide sequence and expression of a mannitol
dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
Escherichia coli
Biochim. Biophys. Acta 1351 (1-2), 157-167 (1997)
97236441
9116029
2 (bases 1 to 10031)
Brunker,P., Altenbuchner,J. and Mattes,R.
Structure and function of the genes involved in mannitol, arabinol
and glucitol utilization from Pseudomonas fluorescens DSM50106
Gene 206 (1), 117-126 (1998)
98121321
9461423
3 (bases 1 to 10031)
Brunker,P., Altenbuchner,J. and Mattes,R.
Direct Submission
Submitted (10-JUN-1997) Institute of Industrial Genetics,
University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany
On Aug 4, 1997 this sequence version replaced gi:2065484.
FEATURES
Location/Qualifiers
1..10031
/organism="Pseudomonas fluorescens"
/strain="DSM50106"
/db_xref="taxon:294"
814..842
932..2242
/gene="mtlE"
932..2242
/gene="mtlF"
/note="similar to E. coli periplasmic maltose binding
protein MaltE; periplasmic subunit of the periplasmic
binding protein dependent mannitol transport system"
/codon_start=1
/transl_table=11
/product="MtlE"
/protein_id="AAC04468.1"
/db_xref="GI:2293414"
TPEASHPKILNWWLEENVRQRLTDTATOGQDFVLTIGMVEAALMGAKGWLEPM
translation="MKFTAKALLACTMTLSAVSLGAQTLTIATVNSDMIRMQLSK
KDLPSYDLDDVFPSPVRSVGLSVKSLYALPFYAESSITVTRDLFKDAGLSMPEHPTW
SQIGFEAKLATDKSKEOYGLCLRGKAGENNALJTLANGYGRWFEKKWQPEENG
EMKDALNFYVDNMKKSPPGASSGNFENLALFNSGKCAIWDASVAGSFVDTKQSK
VADHVGFAPHEKTDKSTWLSYLSLAITSKSKADAKVTSWATSKYGLAVAKT
DGVAANVPFGTRKSTSDYEMKAAPKAVTLESKLVADPTKTLKVPVYIGIQLVTPE
FOAIGTVGKFGSALTGQTVDAALTAQTTEREMKRAGYPK"
2381..3271
/gene="mtlF"
3281..3271
/gene="mtlF"
/note="similar to E. coli MaltF protein; inner membrane
subunit of the periplasmic binding protein dependent
mannitol transport system"
/codon_start=1
/transl_table=11
```


QY	1033	GGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTTGCCCGGAATACCATGGTT	1092
Db	5885	CACAACCCGAGCTGCATGACTGGATCAAGGCCCATGTGAGCTTCCCGAACGCCATGGTC	5944
QY	1093	GACCGATTACGGCTCGTCCGGCAGCAGAACTTCCGGCAGCGATCAAGGCTCAACCGGT	1152
Db	5945	GACCGATACGGCGATGACACAGCAGCCGCCACCGCTGCAACTGCACGATGAACACGG	6004
QY	1153	ATTGCCGATAAAGCGCGGGTAATGGGGAACCTTTATCCAGTGGGTCGGGAAGATAAT	1212
Db	6005	ATCGACGATCCCTGGCCAGGTGTTTGGGAACCCCTTTGTGCAGTGGGTACTGGAAGACAAA	6064
QY	1213	TTCGGTGATCGTCCGTCGGCACCTGGGAAGTCGGTGTTCGAACTGGTGGCGTCGGAATC	1272
Db	6065	TTCGTAACAGCCGCCCGCGGTGGGAAGAGTTGGCGTGCAGTTCACCGACGATGTGACA	6124
QY	1273	CCCTATGAAGAGCGGAAGATTCGCATCTTAACTCTTACACAGTGCATCGCCTGGGCA	1332
Db	6125	CCCTATGAAGAGATGAAGATCGGCTTGCTCAAGCGGACGCCACCTGGCCCTGACCTACCTG	6184
QY	1333	GGTACGTTAATCGTCAAAAATATATCCAGGAAGCAATGACCGATTTATCTATCAG	1392
Db	6185	GGTTTCTCAAGGCTATCGGTTGTGCAGCAGACCATGAACGCCGCTGTTCTCGTGGCC	6244
QY	1393	ATTGCCGACCGCTACGTGACGGAAGATGTCATTCCTTGCTT---GGCGGATAACGGTATC	1449
Db	6245	TACATGCGCGCTACATGGACCTCGAGCTCAGCCAAACCTCGCGCGGTACCGGGCATC	6304
QY	1450	GATTTGCCAACACCTACCGTGATGTGTACTCAAGCGTTTACCAATCCATATTCAGGAC	1509
Db	6305	GACCTGACCGACTACAAGCAGACAGCCCTGGTGGACCGCTTCTCCAACCAAGCGGATTCGCGAC	6364
QY	1510	ACCAACCAACCGCTCGCTCGGATGTTCTCGAAATTCGCGCGGATGATTCGCCCCAC	1568
Db	6365	CAGTTGGAACGGGTGTTCCGATGGCTCGCTCGAAGTTTCCCAAGTTCACCGTCCGCGAC	6423
RESULT	7		
AP003003		349116 bp	DNA linear BCT 15-MAY-2001
LOCUS			
DEFINITION		Mesorhizobium loti DNA, complete genome, section 10/21.	
ACCESSION		AP003003	BA000012
VERSION		AP003003.2	GI:14024067
KEYWORDS			
SOURCE		Mesorhizobium loti (strain:MAFF303099) DNA.	
ORGANISM		Mesorhizobium loti	
REFERENCE			
AUTHORS		Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Mesorhizobium.	
		1 (sites)	
		Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S., Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T., Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A., Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpou,S., Sugimoto,M., Takeuchi,C., Tamada,M. and Tabata,S.	
TITLE		Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti	
JOURNAL		DNA Res. 7 (6), 331-338 (2000)	
MEDLINE		21082930	
PUBLISHED		11214968	
REFERENCE		2 (bases 1 to 349116)	
AUTHORS		Kaneko,T.	
TITLE		Direct Submission	
JOURNAL		Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan	
		(E-mail:kaneko@kazusa.or.jp, URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)	
COMMENT		On May 11, 2001 this sequence version replaced gi:11994978.	
FEATURES		Location/Qualifiers	
source		1..349116	
		/organism="Mesorhizobium loti"	

/strain="MAFF303099"	
/db_xref="taxon:381"	
complement(54..1655)	
/gene="ml13875"	
complement(54..1655)	
/gene="ml13875"	
/codon_start=1	
/transl_table=11	
/product="phosphoglycerate dehydrogenase"	
/protein_id="BAB50672.1"	
/db_xref="GI:14024068"	
/translation="MAPRLVLSVDKLSPTAVQIPKORGVEVDYLDLGKDKKLELVID QYDGLAIRSATKVEKLIINAATRLKIVIRAGIGVDNVDIPAASRKGIIYMNTPFGNSI TTAHAVAMIFALAQIPENASTHAGKWEKRFMGVEITGKTGLVIGCGNIGSVIAT RGVGLMHVAFDPLSDKRAELGVEKVELDELFAADFITLHTPLTKTRNIIDAG ATAKWNGVRIINCARGGLIIEADLIAALKSGKAGAGIDVEVEPAEONALFGEVNV VATPHLGASTAEQNAVLQVAEQMSDYLIKAVSNAINWPSITAEERPLKPFVKLA EYLGAFTGQVTEDEIMEVELFDGSTATMNTALISALAGLIRQVSDVNMVSAPIA VKERGIIVAEVRKDSGVDFGYIKLTVKTEHRTSTAGTCFSDGKPRFIQIKGINLDA EYQHMLYTTNADAPGIIIGLGTGCGENGVIANFOLGRNRPGGDAIALLYLDAPPE KYLEQVRAHKSIDSAKRLQFDVGGI"	
complement(1719..2894)	
/gene="ml13876"	
complement(1719..2894)	
/gene="ml13876"	
/codon_start=1	
/transl_table=11	
/product="phosphoserine aminotransferase"	
/protein_id="BAB50673.1"	
/db_xref="GI:14024069"	
/translation="MTHTKPLGRPANPNFSSGPCAKRFGWSVEALKNAALGRSHRAK IGTKLQEAIELTREILTKVPADHRIGIVPASDTGAVEMALWSLLGSRGVDMAWESFG SGWTDVVKQLKPDVRKEAGYGLPDLAKIDFDRDVTWNGTTSRGVRVNGDFIP ARKGLTICDATSAFAORLDFEKLQVTFESQKVLGGEGAGHMLTILSPRAVARESY KPAWLPKIFRLTSGKLEGIFKGETINTPSCMLCEDYLDLHAKSGTGGDLVLAR ADANAALDGFYDKRSWGLHVAQPATRSNTSVCLSTFDPDVAALDADGQATFAGLV SALDKEGVAYDIGSYRADPPGLRIWCGATVETSDLETLLPLWLDWAPAAQKASLKAAA"	
complement(3096..3869)	
/gene="ml13877"	
complement(3096..3869)	
/gene="ml13877"	
/note="hypothetical protein"	
/codon_start=1	
/transl_table=11	
/protein_id="BAB50674.1"	
/db_xref="GI:14024070"	
/translation="MTLTSRIVLALAGLLPLTAVAADYDPPIVADQAPDVPVEV GSGWYLRGDVSVYLAQSKFNDDPFAITPASDFEKEDEPIFASIGEGYHFNDRYLRADNLG YLPGNKIGIGYDLSVSPFATSTVASADLKNAYSLMUNATVDLGTYYGVTTPYLGCG QYFSAPDAEYVTAASLTSPVHKGISNHQIKLGLRYDLW"	
complement(4111..4944)	
/gene="ml13878"	
complement(4111..4944)	
/gene="ml13878"	
/note="hypothetical protein"	
/codon_start=1	
/transl_table=11	
/protein_id="BAB50675.1"	
/db_xref="GI:14024071"	
/translation="MFTNARKALLALLFAGLASPVFAADLPPEQVEEAPPVVEQRPVD VGGWYIRGDIYHKSDYIDMTYTGIDPCNCTVIPGGSFDYGLKLGSLGGVGY KINDLRADVTADYWFKSNFNGTSDLVSTVSKMSALLLANAYVDIGTWGHTTP YVAGTGGARVKWDTVPDNTAETNPNGASNRFAVALMAGASYCLTDKILIDAGYRFS HIQGGRMFEWDASSAGPGDRCINTHEVRGGLRYQFGGNGNCAAPVVAQPEPEPIYT K"	
complement(5193..6545)	
/gene="ml13879"	
complement(5193..6545)	
/gene="ml13879"	
/codon_start=1	
/transl_table=11	
/product="phosphoglucomutase/phosphomannomutase family	


```
protein: MrsA"
/db_xref="GI:14024072"
/translacion="MAGNYFGTDGIRGRANKFPMTABIAMRVGMAAGLSFQSGSHRH
VWLGKDFRLSSYMTENAMVAGLCAAGMDVLLGPIPTPAVAMLVRSRADIGVMSAS
HNPYDNGIKLFGPDGKLSDEIEERIESMLDKDVELTLADSDGLGRAKRVGDVHDYK
IEFAKRTLRMSLSLGLURIVVDCANGAAYKVAPEALWELGAEEVVAIIEPNFNGINKE
CGSTHPAGLQKHVEVRADIGIALDGDADRVVIYDENGAIVDGQIMAMIAESWHQSG
RLAGGVYSTVMSLQVLEFLGDMKQLQHRTKVGDYVVEHMRAGLNLVGGEGSGHIV
LSDFTTGGLSALQVLACLRQGRVPSLSKFPVPOLLKNVRJAGGKPLPEAPV
KAATEDARNRLGRAGRLVIRPSGTEPLIRVMAEGDDPQLVEAVNDIVEISETRSAA
"
6913. 7779
/gene="mlr3880"
6913. 7779
/gene="mlr3880"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50677.1"
/db_xref="GI:14024073"
/translacion="MLIHRLATLGLAVVTLFLAAPANALTMARCSTKYNAAKDAGTL
AGOTWNOFRACQCTDAATTTDKPATTKAAETPKARATETKPAKAAAAADSTAKGL
TSKCSAKYQAAKAGTLNGMKWDFRTECAACASAAADTKPATTKPATTKAAAP
AGGGLTMAECARYQAQKANTLKGKQWDFRKAEGCAAAADDDSPAPSEANVPE
PAKPTTVPKGVTPPSAISPKYASETEPKGRMHFTCLDQYYALKDANALGLKWTQKG
GFYSLCAKLKG"
8079. 8537
/gene="mlr3881"
8079. 8537
/gene="mlr3881"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50678.1"
/db_xref="GI:14024074"
/translacion="MPNLKLAVLAVTASVAVSAPASALTMKECGEYTRYRAKEGGTLN
GMDAEPKRCATPAAESVSGDVTTPQPKETDAVAPATVAPAGVPTFTLLAAEF
GWKAPAKRKCTQLQYHGKKEAGTLNLRWIKQGGGYSLCNARLKAES"
complement(8624. .10552)
/gene="mlr3882"
complement(8624. .10552)
/gene="mlr3882"
/codon_start=1
/transl_table=11
/product="metalloprotease (cell division protein); FtsH"
/protein_id="BAB50679.1"
/db_xref="GI:14024075"
/translacion="MNEYRNLAUWAIIVLLIALFNLFQTPQTRGASSDVPYSQFLQ
DVAAGRVKTYIAGRTGYTDNSTGFQYSPGDPQLVSLQDKNVTINARPEADGS
NSLFGYLISLPMILILGWIFFMROMQSGSRAMGFKSKAKILLTEAHGRVTPQDVA
GVDAKEDLEEIVELFPQKFORLGGKIPRGVLLVGGPGTGKTLARSAVAGEANVP
FTISGSDVEMFVGVGASRVDMEDQAKNAPCIIFIDEIDAVGHRGAGLGGGNDER
EQTUNQLLVENDGESNESIILIAATNRPDVLDPALNPRFRDQVVPNDIYGREK
ILKVHVRNPLAPNLDLVKVARGTFPGFSGADLMNLVNESALMAARNRKLRLVMAEFED
AKKIMGAERSAMTQAQKELTAYHEAGHAILALNVPSSADPLHKATITPRGRALGH
VMQLPEDGRSMYSKYMSRLAIIMMGVRVAEPEFKENITSGASSDIEQATKLARAM
VTRWGFSDKLGHVAGDNOEIVFLGHSVARTONISEETAQIIDEAVERLIDEAYSTAK
SILTKKKKHWALAQGLLEYETLSGDEIKQLIAGEKPARDLGDDTPSRGSVAPKSGG
RRKGPPEGMPEQPS"
complement(10675. .12039)
/gene="mlr3884"
complement(10675. .12039)
/gene="mlr3884"
/note="Similar to YaeN protein [Bradyrhizobium japonicum],
contains similarity to cell cycle protein MesJ/cytosine
deaminase-related protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50680.1"
/db_xref="GI:14024076"
/translacion="MLDTEPDLSTRLFSHIDFTHGAVAAVSGSDSTALLLLLHHFED
RTGSSAKLLAVTIDHGLRQSAEAQAVAKLCAERGAHRLTLTWTRKRPSTGLPAAAR
```

```
DARYHLLAEARAEIGLIVTGHHTADDOAETVLMRHRDRELADLAGRLAGMAPATL
YDREWIVRPLLTGRRSALRDFLRREHVWAEADPTNIDEAFPRVRAALAGSAGHM
ENTLRLAQAAVERGQLGASAAANIRRIASQPSSTGLRLDLPALLIADQAAIYALRIL
LATAGGAFLPDQASAAFLVRLKAGFLCATLSRTVDFRHAGFLLRREARGLPPAAT
AVDNTINDGRRHITLNDMSGALLTAPLGAAARLALDDGETPASLLIRAAALAEPTLL
QAFENLGPQSEPLPVEAARPVVSPFARFLPSFDLAPARAVAEIIGNSLPASPLGG
HSAO"
complement(12098. .13210)
/gene="mlr3886"
complement(12098. .13210)
/gene="mlr3886"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50681.1"
/db_xref="GI:14024077"
/translacion="MHLRSVLSGTLAALLLSGFTALASGPGASGPGSTDSFSFHLPL
SLELPSLFGDKKPDQVQFAQOSNGAGATGLEDLROMNGKIEELNFOILQOMRMRK
QOEDNEFRFOLEGGAGGQPPKPKSDATNTNTDVAANPATQAPADAGAAPGNQS
AGKTEFEDVIVESPEGDPKVPFGTGAPEKFTGSITVDKNGNVIDAGANTQSTAPTD
AAPATKAPAKAGKNGTIVAAPLSTNDPEELYRNSYQFILSGDYGTABQGRDRHSRF
"
Query Match 6.1%; Score 113.2; DB 1; Length 349116;
Best Local Similarity 46.9%; Pred. No. 3.5e-23;
Matches 524; Conservative 0; Mismatches 558; Indels 36; Gaps 4;
Oy 503 TGCACATCGGGTTAGTCTCTTTTCATCGCGCACATCAGCGGTGGTATCTACACCGTTTCG 562
Db 132260 TGCATCTCGCGTCGCGCCCTTCCATCGCGCCCATCAGCGCCCTATGTGCGACGAGTGCC 132319
Oy 563 AGGTGATGGCGATAAACGCTGGAGCATTGCTGCGGGCAATATTCGTAATGATGCTGAAC 622
Db 132320 TGGCGCGGGCGAAACGAGCTGGGCATACAGGGGCTCTCCTTGGCGACCCCGCACACG 132379
Oy 623 ATGTGCTACAGCACTAGTGCACAGAAAGTGCCTATGTCTGTGAAACCGTCACCCCGG 682
Db 132380 GCGATCGCGTGTGCG-----CGCAGAGCGGCTCTACACGTGGGATCCCGACGAGTG 132433
Oy 683 AAGGGTAAAGCAATATGAAGAGATCACCTCAATTCAGAAGTTGATACCGTGGCAGGACG 742
Db 132434 GCGAGGA---GGAGCTGTGCGTCATCGGCTCTATCCAGTCGATGCTGCTGGCGCGGAAA 132490
Oy 743 ATTTACAACCGCTGATTGCTGAAGGGCGAGATCCGAAGACAAAGATGATGCTTTCACCG 802
Db 132491 ACCCCAGCGCGGTGTCGCGCTGACCGATCCACGACCCCGCATCGTCACGCTGACCA 132550
Oy 803 TCACCGAAAGCGGTACTACCTGAATACCACTACCAAACTGGAAGTTAAACAATCCTGATT 862
Db 132551 TCACCGAAAGCGCTATCTCAGGGCTGCGGGCGCGGATGGACGCCGCCCATCCGACA 132610
Oy 863 TAGCGGCAGATCT-----TAAAGGGGGATGCAAAACAAATTTACGCTGTTATTACCCGTA 916
Db 132611 TCATCCAGCATTGCCCAATCCGCAATCCCGAGGCGCGCATGGTTTCTTCGCGGAGG 132670
Oy 917 TCCTCGAAGCGGTATGGAATAATACGCGGACCACTAACCCTGCTGAATTCGGATAACG 976
Db 132671 CGCTTTCGCGCAGACTGGCGCGCGCACCCGCCCTTCAACCGTGTGCTGCGACAACC 132730
Oy 977 TCGGCCAATAAGTGAACGTTTCCATGATGGCTGTTGAGTTTCTCCAGCTA----- 1029
Db 132731 TCGCGGCCAACGGCGCCACGCTGCATCGGCTGTCTACCGCAATTCGCAACATTCGCGGATG 132790
Oy 1030 -----ACTGGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCAATACCAC 1075
Db 132791 CGAGTCTACCACTGGAGCGACACAGNAATCCCGACCATATCGCGAACCAAGGTCCGCT 132850
Oy 1076 GCCCGAATACCATGTTGACCGCATTACGCTCTGTCGGCAGCAGAACTTCGCGGACGGA 1135
Db 132851 TTCCGTCGAGCATGTCATCGCATCGTGCCTGCGCAACCAACCGACCGCGGCGCGCA 132910
Oy 1136 TCAAGGCTCAACAGGGTATTTGCCGATAAGCCGGTAAATGGCGGCAAACTTTATCCAGT 1195
Db 132911 TCGCTGGGAACTCGTGTGCGACGATGCTGGCTGTCTATGATGACCGAGCCCTTCCGCAAT 132970
```

QY 1196	GGGTGCTGGAAGATAAATTTCCGTGATGTCCTCGGCACTGGAGAAGTTCGGTGTGGAAC	1255
Db 132971	GGGTGCTGGAAGACAATTTCTCCGCGGAGCGGAGTGGGAAAAGTTTCGGCGCTCAACA	133030
QY 1256	TGGTGGCGTGGTAATCCCTATGAGAGCGGGAAGATTCGCATCTTAACCTCTTACACA	1315
Db 133031	TGGTGGCGATGTCGGCGCTTCGAGGACATGAAGCTCAGGCTGCTCAACGGCGGCATT	133090
QY 1316	GTTCGATCGCCTGGCGAGGTACGTTAATCGGTCAAAAATATATCCAGAAAGCACATGA	1375
Db 133091	CGGCGATCGCCTATCTCGGCGCTGCTCAGCGGCCACCGCCACCGTCGACCGCGCTT	133150
QY 1376	CCGATTTATCTATCAGATTTCGGACCGCTACGTGACGGAAGATGTCATTCCTTGTCTG	1435
Db 133151	ATCGGGGATCCGGCAATTTGTCAGCGGCTCTGGCGGAGGCCATCCCGACATTCGCGG	133210
QY 1436	GCATACGCTATCGATTTCGCAACCTACCGTGATGTTGTACTCAAGCGTTTACCAATC	1495
Db 133211	AGGATCGCGACTCGATACGGAACCTACACAGCGAGCTTCCGCGAGCTTCTCCAACA	133270
QY 1496	CACATATTCAGGACACCAACCAACCGCTCGCTCGCGATGTTTCTCGAAAATTCGGGGA	1555
Db 133271	CCGCGTTGCCACCGCAGCGGCGAGATGCCAATGACGCGACCGCAGAAAGTCCGCGAGC	133330
QY 1556	TGATTGCCCCACACTGCGAGAGTGTACACGCGAGGC	1593
Db 133331	GCATCATCGGCTCGGCGATCGAGTGCTCGAAGCCGCGC	133368
RESULT 8		
SCYNR073C/c		
LOCUS	SCYNR073C	3250 bp DNA linear PLN 11-AUG-1997
DEFINITION	S.cerevisiae chromosome XIV reading frame ORF YNR073c.	
ACCESSION	271688	Y13139
VERSION	271688.1	GI:1302609
KEYWORDS	Saccharomyces cerevisiae.	
SOURCE	Saccharomyces cerevisiae	
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	
REFERENCE	1 (bases 1 to 3250)	
AUTHORS	Andre, B., Iraqui Houssaini, I., Urrestarazu, L.A. and Vissers, S.	
JOURNAL	Unpublished	
REFERENCE	2 (bases 634 to 3250)	
AUTHORS	Levesque, H., Lepingle, A., Nicaud, J.M. and Gaillardin, C.	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 3250)	
AUTHORS	MIPS.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-APR-1996) Data collected by MIPS on behalf of the European yeast chromosome XIV sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Newes@mips.embnnet.org	
FEATURES		
source	1. 3250 Location/Qualifiers	
	/organism="Saccharomyces cerevisiae"	
	/db_xref="taxon:4932"	
	/chromosome="XIV"	
	complement(441..1949)	
	/note="ORF YNR073c"	
	/codon_start=1	
	/protein_id="CAA96356.1"	
	/db_xref="GI:1302610"	
	/db_xref="SWISS-PROT:P39941"	
	/translation="WTKSDFTATSIKAKTLKSFESTLPIPTYPREGVKQGVHLGVG	
	AFRSHLAVFMRHLKQEHKLDKNSICGVLKMDALMDKAAQOCLTLVBRGNDT	
	NATVGSITAYMAFDPRVTERKANPDTHIVSTVTEGHHYHSNSLMDTDAPEI	
	INDLHNEPKDPTLYLYEALLRYKRLPTIMSCNDMPQVTVTMVAFKLK	
	KDEKFAEDVTSPNSMDRVTPRCDTKERYVADTWGIDKQCPVVAEPFIQWLE	
	DNFSDGRPPWELGVQVVKDVSLEMLKRLNLLGGHSAMGYLGAGTYIHEVNDP	
	TINKYIRVLMREVIPLLPKVPGVDFEYTSVLFRESNPATQDVARICLMSGKMP	
	KVLPSIYQLRKPDKYKLLAVCVAGHFRILYTGVDNMCKPEIEDHPAPTLLKAAVK	

BASE COUNT	968 a	651 c	638 g	993 t
ORIGIN				
Query Match	6.1%	Score 113;	DB 8;	Length 3250;
Best Local Similarity	49.8%	Pred. No. 1.6e-23;		
Matches 315;	Conservative 0;	Mismatches 315;	Indels 3;	Gaps 1;
QY 949	CCACTAACCTCTGCTGAATTCGGATAACGTGCGGCCATAATGTGAACCTTTCCATGATGGC	1008		
Db 1379	CCATTCACCTATTATGTCATGTGACAACATGCCCAAAATGGTGTACAGTAAGAACCATG	1320		
QY 1009	CTGGTGTAGTTTCTCCAGCTAACTGGCAACAGGATGTCATCGACTGGCTGAGTACAAAT	1068		
Db 1319	CTTTGTGCAATTTTGGCAAGTAAAGAAGATGAGAAATTCGCGCGCTGGATTTGAAGACAAG	1260		
QY 1069	ACCACTTGGCCGAATACCATGTTGACCGCATTAAGCCCTCGTCGCGGACGAGAATTTCCG	1128		
Db 1259	GTTACTTCTCCTAACAGCATGTGTGACCCACCTGTGTACCCACCTGTGTACCGATAAAGCGGT	1200		
QY 1129	GCACGGATCAAGGCTCAACACGGGTATTCCGGATAAAGCCGCGGTAAATGGGCGAACCCTTT	1188		
Db 1199	AAATAGTTTGTGACACCTTGGGGAATCAAGGATCAATGTCCCGTTGTCGCAAGACCTTTC	1140		
QY 1189	ATCCAGTGGGTGCTGGAAGATAAATTCGCTGATGTCTCCGTCGCGCACTGGAGAAGTCGGT	1248		
Db 1139	ATCCAATGGTCTTGAAGACAACCTTCTCCGATGCGCGCTCCTCCATGGGAATTTGTTGGT	1080		
QY 1249	GTCGAAGTGGTGGTCGGTAATCCCTATGAAGAGGGGAAGATTCGATCTCTTAACCTCT	1308		
Db 1079	GTTCAAGTCTCAAGGATGTCGATTCTTACGAATTGATGAAGTTGCGCTCTACTTTAAGCGGT	1020		
QY 1309	TCACACAGTTGTCATCCCTGGGCGAGTACGTTAAATCGGTCAAAAATATATCCACGAAGC	1368		
Db 1019	GGACATCTGCTATGGGATATTGGGATACATTGGCAGGCTACACCTATATACATGAGGTT	960		
QY 1369	ACAATGACCGATTTTATCTATCAGATTGCCACCGCTACGTGACGGAAGATGTCTATTCCT	1428		
Db 959	GTCACAGACCCCAACTATCAACAAGTATATCCGTGTTTGTGTCGCTGAGGAAGTTATCCCA	900		
QY 1429	TGCTTG---GGCGATAAAGGATATCGATTTGCCAACCTACCGCTGATGTTGTACTCAACGCT	1485		
Db 899	TTATGCTCTAAAGTCCAGGTGTTGATTTGAGAGTACATGCTATCATGTTCTTGGGAAGA	840		
QY 1486	TTTACAATCCACATATTTCAGGACACCAACCAACGCGTCGCTGCGGATGTTTCTCGAAA	1545		
Db 839	TTCTCCAATCCAGCTATTTCAGGACACTGTCGACAGTATTGTTGATGGGCTCTGTTAAG	780		
QY 1546	ATTCGGCGATGATTGCCCCACACTGCGAGAG	1578		
Db 779	ATGCTCAAGTATGTTTTCATCAATTTACGAG	747		
RESULT 9				
SCYNR074C/c				
LOCUS	SCYNR074C	4931 bp DNA linear	PLN 11-AUG-1997	
DEFINITION	S.cerevisiae chromosome XIV reading frame ORF YNR074c.			
ACCESSION	271689	Y13139		
VERSION	271689.1	GI:1302611		
KEYWORDS	Saccharomyces cerevisiae.			
SOURCE	Saccharomyces cerevisiae			
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
REFERENCE	1 (bases 1 to 4931)			
AUTHORS	Andre, B., Iraqui Houssaini, I., Urrestarazu, L.A. and Vissers, S.			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 4931)			
AUTHORS	Levesque, H., Lepingle, A., Nicaud, J.M. and Gaillardin, C.			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 4931)			
AUTHORS	MIPS.			
TITLE	Direct Submission			

JOURNAL Submitted (29-APR-1996) Data collected by MIPS on behalf of the European yeast chromosome XIV sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mews@mips.embnet.org

FEATURES

source Location/Qualifiers

1..4931

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="XIV"

complement(<1..1316)

/note="ORF YNR073c"

/codon_start=1

/protein_id="CAA96358.1"

/db_xref="GI:2253199"

/db_xref="SWISS-PROT:P39941"

/translation="MTKSDTTATSLNAKTLKSPSTLPITYPREGVKQGIHVLGVG AFHSHLAVFMHRLMOEHLKDWISCGVLMKADALMDAMKAQDCILYVERGIKDT NAYTVGSITAYMAYAPDDPRAVIEKMANPDTHIVSLVTENGYHSEATNSLMDAPEI INDLNHPKEDTLGYLYEALLRYKRLGTPFTIMSCDNMPQNGVTVKTLVAFAKL KDKFAAIEDKVTSPNSMVDVPTPCRTDKERYVADTWGDKDCPVVAEPFIOWVLE DNFSDGRPPHELGVQVVDVSYELMKRLNGHGSAMGYLGYLGYTYIHEVNDP TINKYIRVLMREEVIVLLPKVPVDFEETASVLERFSNPAIODTVARICLMGSGKMP KYVLPSYEOLRKPDGKYKLLAVCVAGWERYLTGVDNMNGKPFIEDP"

complement(2618..3754)

/note="ORF YNR074c"

/codon_start=1

/protein_id="CAA96357.1"

/db_xref="GI:1302612"

/db_xref="SWISS-PROT:P52923"

/translation="MTINTKNIVVVGAGVFGVSVANHLYRELGYTAIKLVTSNYYV FLPSAVRLTVSKDTSYKILPLKNVLDGSEIVIKDTAASFDKKEVVLGSDRAIKPDLV LATGSKWADPIGTYTGDNYKEYFEREASISDADHILFLGGFVNCLEAGELLFVY LEETRSGKKRISIIHNSDKLLPDSGLYNLKNVTDYLSKNGITLVNTVGSALDTS PKRFLGEGSSKYIDADLYRGVGSIPNPVNSISDLCKDKGFTQVEKNFRVAVEAG NVAIGDVTNFRHGLVKRONVDVLDVTRNVISSLOEGTEASLDADCLTGHAPSQVSGV LGPNAGRGQFPLPLGLTINIPSLFISIRAKSNLFSDRMEPLFKK"

BASE COUNT 1531 a 1036 c 900 g 1464 t

ORIGIN

Query Match 6.1%; Score 113; DB 8; Length 4931;

Best Local Similarity 49.8%; Pred. No. 1.7e-23;

Matches 315; Conservative 0; Mismatches 315; Indels 3; Gaps 1;

Qy 949 CCACCTAACCCCTGCTGAATTGGCAATACGTCGCCATAATGGTGAACGTTCCATGATGCC 1008

Db 746 CCATTCACATATTATGTCATGTGACAAACATGCCCAAAATGGTGTACAGTAAGACCATG 687

Qy 1009 CTGGTTGATTTCTCCACCTAACTGGGCAACAGGATGTCATCGACTGCTGAGTACAAT 1068

Db 686 CTGTGTGATTTGCCAAGTTAAAGAAGGATGAGAAATTCGCCGCTGGATTGAAGACAAG 627

Qy 1069 ACCACTTCCCGCAATACCATGGTTGACCGCATACGCCCTCGTCGGCAGCAGCACTTCG 1128

Db 626 GTTACTTCTCTACAGCATGGTGGACCGTGTGACCCACAGTTGTACCGATTAAGAGCGT 567

Qy 1129 GCACGGATCAAGCGGTCAACCGGTTATTCGCGATAAGCCCGGTAATGGCGGAACACCTTT 1188

Db 566 AAATACGTTGTGACACCTGGGAATCAAGTCAATGTCCTGGTTGTCGCAGAACCTTTC 507

Qy 1189 ATCCAGTGGTGGTGGGAAGATAATTTCCGTCATGTCCTCGGCACTGGAGAAGTCGGT 1248

Db 506 ATCCAATGGGTCTTGAAGACAACCTTCCTCCGATGGCCCTCTCCATCGGGAACCTTTGGT 447

Qy 1249 GTCGAACGTGGTGGCTCGGCTAATCCCTATGAAGAGCGAAGATTCGCATCTTAACCTCT 1308

Db 446 GTTCAGTCTCAAGGATGTGATTCCTACGAATGTGAGAGTTGCGCTACTCTTAACCGT 387

Qy 1309 TCACACAGTTGCATCGCTGGCGAGGTACGTTAATTCGGTCAAAAATATATACACGAAAGC 1368

Db 386 GGACATCTGCTATGGGATATTGGGATACCTTGGCAGGCTACACCTATATACATGAGTT 327

Qy 1369 ACAATGACCGGATTTATCTATCAGATTCCGACCGCCTACGTCGACGGAAGATGTCATCCT 1428

Db 326 GTCAACGACCAACATATCAACAAGTATATCCGTGTTTGTGTCGGTGAGAAAGTTATCCCA 267

Qy 1429 TGCCTTG---GGCGATAACCGGTATCGATTTCGCAACCTACCGTGATGTTGTACTCAAGCGT 1485

Db 266 TTATTGCCATAAGTCCAGGTGTTGATTTGGAAGAGTACACTGCATCATGTTGGAAGA 207

Qy 1486 TTTACCAATCACATATTCAGGACACCAACCAACGCGTGGTGGGATGGTTCCTCGAAA 1545

Db 206 TTTCCCAATCCAGCTATTCAGGACACTGTCGCACGATTTTGTGATGGCTCTGCTAAG 147

Qy 1546 ATTCGGCGGATGTTGCCCCACACTGCGAGAG 1578

Db 146 ATGCTTAAGTATGTTTTCGCCATCAATTTACGAG 114

RESULT 10

SCS288C14/C SCS288C14 9172 bp DNA linear PLN 27-NOV-1995

LOCUS SCS288C14

DEFINITION S.cerevisiae N3810, N3815, N3820 and N3825 genes.

ACCESSION X86790

VERSION X86790.1 GI:805055

KEYWORDS S288C.

SOURCE Saccharomyces cerevisiae.

ORGANISM Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 9172)

AUTHORS Levesque,H., Nicaud,J.M., Lepingle,A. and Gaillardin,C.

TITLE Sequence of a 9.2kb telomeric fragment from the right arm of S. cerevisiae chromosome XIV

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 9172)

AUTHORS Nicaud,J.M.

TITLE Direct Submission

JOURNAL Submitted (02-MAY-1995) J.M. Nicaud, Institut National de la Recherche Agron., Genetique Moleculaire et Cellulaire, Thiverval-Grignon, F- 78850 Thiverval Grignon, FRANCE

FEATURES

Location/Qualifiers

1..9172

/organism="Saccharomyces cerevisiae"

/strain="S288C"

/db_xref="taxon:4932"

/chromosome="XIV"

/map="right telomere"

/clone="pEL186B"

complement(1..1316)

/gene="N3810"

complement(<1..1316)

/gene="N3810"

/codon_start=1

/protein_id="CAA60486.1"

/db_xref="GI:805056"

/db_xref="SWISS-PROT:P39941"

/translation="MTKSDTTATSLNAKTLKSPSTLPITYPREGVKQGIHVLGVG AFHSHLAVFMHRLMOEHLKDWISCGVLMKADALMDAMKAQDCILYVERGIKDT NAYTVGSITAYMAYAPDDPRAVIEKMANPDTHIVSLVTENGYHSEATNSLMDAPEI INDLNHPKEDTLGYLYEALLRYKRLGTPFTIMSCDNMPQNGVTVKTLVAFAKL KDKFAAIEDKVTSPNSMVDVPTPCRTDKERYVADTWGDKDCPVVAEPFIOWVLE DNFSDGRPPHELGVQVVDVSYELMKRLNGHGSAMGYLGYLGYTYIHEVNDP TINKYIRVLMREEVIVLLPKVPVDFEETASVLERFSNPAIODTVARICLMGSGKMP KYVLPSYEOLRKPDGKYKLLAVCVAGWERYLTGVDNMNGKPFIEDP"

complement(2618..3754)

/gene="N3815"

complement(2618..3754)

/gene="N3815"

/codon_start=1

/protein_id="CAA60487.1"

/db_xref="GI:805057"

/db_xref="SWISS-PROT:P52923"

/translation="MTINTKNIVVVGAGVFGVSVANHLYRELGYTAIKLVTSNYYV FLPSAVRLTVSKDTSYKILPLKNVLDGSEIVIKDTAASFDKKEVVLGSDRAIKPDLV LATGSKWADPIGTYTGDNYKEYFEREASISDADHILFLGGFVNCLEAGELLFVY LEETRSGKKRISIIHNSDKLLPDSGLYNLKNVTDYLSKNGITLVNTVGSALDTS

[illegible]

Df 206 TTCTCCAATCCAGCATTTTCAGGACACTGTGCACGTAATTGGTGTGAATGGCGCTCCTGGTAAG 147

Gy 1546 ATTCCGCCGATGATTGCCGCCCACTCGGAG 1578

Df 146 ATGCCTAAGTATGTTTTGCCATCAATTTACGAG 114

RESULT 11
AE009374

LOCUS linear BCT 20-DEC-2001

DEFINITION Agrobacterium tumefaciens str. C58 (U. Washington).

ACCESSION AE009374.AE008689

VERSION GI:17742928

KEYWORDS Agrobacterium tumefaciens str. C58 (U. Washington).

SOURCE Agrobacterium tumefaciens str. C58 (U. Washington).
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.

REFERENCE 1 (bases 1 to 11164)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Boeve Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McCllland,E., Palmer,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim.S., Hendrick,C.,
Zhao.Z., Dolan.M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.

TITLE The genome of the natural genetic engineer Agrobacterium
tumefaciens C58

JOURNAL Science 294 (5550), 2317-2323 (2001)

MEDLINE 21608550

PUBMED 11743193

REFERENCE 1 2 (bases 1 to 11164)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Boeve Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McCllland,E., Palmer,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim.S., Hendrick,C.,
Zhao.Z., Dolan.M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.

TITLE Direct Submission

JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA

FEATURES source location/Qualifiers
1..11164 /organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
94..1578 /gene="mtk"
/note="Atu4451"
94..1578 /genes="mtk"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="mannitol 2-dehydrogenase"
/protein_id="AAI45245.1"
/translation="WTCKLSLATLDEAKKYAAVPAYSRNDLSAGIVHFGVGNFHRAHQ
AVYLDOLFNTGDADHPAIIGCVLPDSATMREKLAAODELTITVEQDNRTGARVTGP
MDIKVFQLAITDLAPDKIRIVMTITEGGFYIDSAGSNPQHOPATAEKDGKNFS
PTLVCFGI VAGLKARDKDLOPFTMSCNDIPHNKGKVTKNAVVGSLALSDPAEFANNIC

```

ENVAFPNSMVDRTTPATGERERNIARDGIEDNWPVCEEFKOWNMEDNFPAGRPAI
EKAGVQVKVAPYEHMKIRILANGHAAIAYPAALLDIHVFHEAMEHPLIRLAKLE
KEGIPVIPPVOTDLNDYFGLERFLPNKIGDTIPRLAQDGSNROPKTLILSTLDR
LSRGEIVGLSLVAKMCRFYGTSDSGKEIVENDASAERLOAAAIRAKDDPVAFLAL
DDLFGVAESALFRKFSHALTILWQEGTAKTILQYLDLDDLAEK"
1585. .2304
/gene="Atu4452"
1585. .2304
/gene="Atu4452"
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45246.1"
/db_xref="GI:17742930"
/transl_table="MRDGTGPLLIFDCDGLVDSEVPYSIAVLLDMLSHLGVTWGEEEA
YERFLGRSVASMTTLFEETGVETDIDFLDHMRATLFRERTELKPIDGIAETLDELL
PLKRCVASSQPERIRYSGLTGLIDKFEPHFVSATVVRNGKPAPOFLFLHAARSMGYE
PROCVIEDSPAGIAAKAAGMGVFAFTGGSHARFPFAFREKIAIGLADAVFDAMPDLV
QLVGSYVGRGGDQOVERDAG"
2359. .3936
/gene="Atu4453"
2359. .3936
/gene="Atu4453"
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ribitol kinase"
/protein_id="AAL45247.1"
/db_xref="GI:17742931"
/transl_table="MMRONLVADVGVTSARAGIFDPAGRLIARSIHPILMOREPNEH
AENDSTIWNACIYKAAALADAGVLQSTAAIGFDATCSLYIRDERGEVPSYSTDD
DRDRTIWLDRHRAIGEADRLTASCHRVLDFAGNSVSEPMQMPKMLWLTTHMPSWSRM
PFTFDLADLTWATKATGSAQSNCTQTAKWNFLAQEHQPGQADYLAFAGLDLDERAGL
SEFVMPGGAIGPLSPAAEALGLDTCQVAAGMIDYAGALGALGCLAEADVDGKHVA
LIAGTSSCLVAMSTQPMGRSLMGYPYQWAVLPGLHWLVEGGQSATGALLDHIVRMHAAQ
GEPTALHARIVARVITBELREGEAFADRLHLVPDFHGNRSPLADPHAVGVVSLTLD
TSFDSLCLRYWRTAIALGARIHVLDAMERFGYAVETLHVTGGHVKNPLLMELYADVT
KRIVVPATADAVLLGTAMTAATAGVHASLAAGAAMYPGNAEISNPALAAHYERD
YRFLANRYRHOELES"
complement(4041. .4874)
/gene="Atu4454"
complement(4041. .4874)
/gene="Atu4454"
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Rpir family"
/protein_id="AAL45248.1"
/db_xref="GI:17742932"
/transl_table="MVEKLLNLTRAMKAGTPSERKIAKYLIEHLDELPETAQTLLAA
KLSPSTVGRFLRSLGYRQFSDIRADLRHAEETVAGDQPVTAEGGRHSNPSQLLS
QQOIAQTADMTQTTLWRAMNEITSADVCLATSPGLDAGRHFGYRLLECGRHIIH
YLGSDNAYVALWDYDPARTLLVIMDCGNLLPLQRLSTTARKGGYRTLLITTFYEW
GPESVDLCLAMPQSONGGQGLQLVLSLEFTLCALSAGADDAAGKARVKYNLAALKRSLR
S"
4897. .5067
/gene="Atu4455"
4897. .5067
/gene="Atu4455"
/located using Glimmer"
/located using Glimmer"
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL45249.1"
/db_xref="GI:17742933"
/transl_table="MNGDAFARPCCKRANHDI.TVETSPFGHVLHRTGMRIIVDGLLYFR
AIFYSRIPFPAV"
5170. .5907
/gene="Atu4456"

```

```

5170. .5907
/gene="Atu4456"
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45250.1"
/db_xref="GI:17742934"
/transl_table="MIFDAARLAFGLNFAAETRSVFWKVLGLTLVLAAALWFAIRSF
IYFALPWIDILIPGVPDWAGNLSFVFAIFAGIGLALALALISPTAVIAGLFLDDVA
DVEKQNYDDPPGKAMPIGEAVLUSSIRKFGVWJAGNLVALLLLVPGVNLIAEFLVN
GYLIGREFFFAAMRFRSPADARLFRYKRTTVPFAGLVIAAFVLAFLVPELNLTLPLFAA
SLMVHLHKAVSRDRDPSPAAGTEQLRG"
complement(5869. .6579)
/gene="Atu4457"
complement(5869. .6579)
/gene="Atu4457"
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="endonuclease III"
/protein_id="AAL45251.1"
/db_xref="GI:17742935"
/transl_table="MRVKATYSKDELTEIFRRFSIQRPPEKGELEHTNPFTLLVAVAL
SAQATDGVNARATRALPKFVADTPKMLALGEEQLIGHIKTIGLYRNKAKNYIALSOML
IDNKGVPKTRREELVTLPGVGRKTANVMSMAFGVPTLAVDTHVFRANRCLAPGK
TTDEVEDRLVRIIPQELFLFAHHLWLIHGRYCCARKRPECERCVIADICKSPKFTFI
PAPILVELPAQLFGAAGGE"
6689. .7171
/gene="Atu4458"
6689. .7171
/gene="Atu4458"
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL45252.1"
/db_xref="GI:17742936"
/transl_table="MDTLNDQPIFAELVPHRSIGRGFRULLLITGPAFLIYGGFFL
ATGAWPIGLFGLDFLLYIAFRANYRAAKAEVSVSRNLSTKRKSPAGRMVYERF
NPWEARFVRHRHDEBIGVSMHVTEGRATDIGSLFNPDDRESFAKAFGALATVRRRM
"
7438. .8310
/gene="ada"
7438. .8310
/located using Blastx/Glimmer"
/located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="6-O-methylguanine-DNA methyltransferase"
/protein_id="AAL45253.1"
/db_xref="GI:17742937"
/transl_table="MNANIMLNEDITPTIGSDYDTRVGVIELTLLDYREQPSLEATAAR
LGQSPQLOKQFTFRWAGLSPKAFLOAVTLDHAKRLRREEDLPLETSEVMSGPGRL
HDLFVTHEAMSPGEMKARGGGLTIRYGFHASPFGALVMTDRGLAGCAFDAPGDERA
CFEDMAGRWNPNADVEDREATYAAARIFEPAMWTADKPLRVLLGDTQFVRVWKSLL
KIPMGRAVTYSNIACIDICQPTASRAVGAAGVNPVSVFVVPCHRAVKGSGALTGYHWGL
TKKRAMLGWETGKA"
8375. .8599

```

Query Match 6.1%; Score 112; DB 1; Length 11164;
 Best Local Similarity 48.1%; Pred. No. 4.2e-23;
 Matches 415; Conservative 0; Mismatches 435; Indels 12; Gaps 3;

Oy 742 GATTTACACCGCTGATTCGTAAGGGCAGATCCGACAAAGTGAATTCCTTACC 801
 ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 424 GATACCCAGGGCATCATCATGATACGTTCGCCGATCCGAAANTCCGATCGTTTCGATGACG 483

QY	802	GTCAACGAGCGGGTACTACCTGAATACACAGTCAACAACCTGGAAGTTAACAACTCCTGAT	861
Db	484	ATTACCGAGCGGGTATTATTCATCGATCGCTCGCGTCTCTTCAATCCGACGACATCGGGC	543
QY	862	TTAGCGGCAGATCTTAAGGGGGATGC-----AAACAAATTTACGGTGTATTACCCCGT	915
Db	544	ATTGCGGAAGACGGCAAGAACCGCTCGGTCCGAAACCGTGTTCGGTGTTCATCGTGTGCC	603
QY	916	ATCCTCGAAGCGGTATGGCAATAACGCGGACCACTAACCCCTGCTGAAATTCGGATAAC	975
Db	604	GGGCTGAAGCGCGCGCGATAGGACCTTGCACGCCCTTCACCGTCATGTCTCGACAAAC	663
QY	976	GTGCGGCATTAATCGTGAACGTTTCCATGATGGCCCTGGTGTGAGTTTCTCCAGCTAACTGGC	1035
Db	664	ATTCCCAACAATGGCAAGGTGACGAAACCGCGTGTGGGCTGGCAGCGCTCTC---C	720
QY	1036	AAACAGATGTCATCGACTGGCTGAGTACAAATACCACTTGCCCGCAATACCAGTGGTTGAC	1095
Db	721	GACCGGCCTTTCCGAACCTGGATCGGCGAAATGTGCGCTTCCGAATTCATGTTGTCGAC	780
QY	1096	CGCATTACGGCTCGTCCGGCAGCAGAACTTCCGGCACGGATCAAGGCTCAAAACGGGTATT	1155
Db	781	CGCATCACCCGCCACCGCGAGCGGTGAACGCAATATCGCCGTCGACGATTTTCGGCATC	840
QY	1156	GCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCAGTGGTCTGGAAGATAATTTTC	1215
Db	841	GAGGATAATTTGGCGGTCTTCTCGGAAGAGTTTCAAGCAATGGGTGATGGAGGATAATTTTC	900
QY	1216	CGTGATGTCCGTCGGCACCTGGAAGAGTCCGTGTGCGAACTGGTGGCGTGAATCCGCC	1275
Db	901	CCGCGAGCGCTCCGGGCTCGAAGAAGCGGGGCTCCAGTTCGTCGAAGATGTTGCGGCC	960
QY	1276	TATGAGAGCGGAAGATTCGCACTTTAACTCTTCACACAGTTGTCATCGCCTGGGCGAGTT	1335
Db	961	TACGAGCACATGAAGATCCGCATCCTGAATGGTGGCCACGCGCAATCGCCTATCCGGCA	1020
QY	1336	ACGTTAATCGGTCAAAATATATCCACGAAGACAGAAATGACCGATTTTATCTATCATGATT	1395
Db	1021	CGCTGCTCGACATTCATTTCGTGTCATGAGCGGATGGAGCATCCGCTGATCCGGCGGTTT	1080
QY	1396	GCGGACGCTACGTACGAGGAAGATGTCATTCTTCGTCTT---GGCGCATACGGTATCGAT	1452
Db	1081	CTGCGCAAGCTCAAAAAGACGAGATCATCCCGGTTCATTCGCGCGGTACTTGACACCGCA	1140
QY	1453	TTGCCAACCTACCGTGATGTTGCTACTCAAGCGGTTTACCAATCCACATATTCAGGACACC	1512
Db	1141	CTCAACGACTATTTTCGGGCTGATCGAGCGCGCTTTCCTCAACCCGAAGATTGCGCATACC	1200
QY	1513	AACCAACGGCTCGCTCGGATGTTTCTCGAAATTCGCGGATGATTCGCCCCACACATG	1572
Db	1201	ATTCCGCGCTCGCGAGGATGGTTCACACCGCCGAAATTCATCTCTGCGCCTCCACT	1260
QY	1573	CGAGAGTGCTACACGAGCGGCG	1594
Db	1261	TTAGATCGGCTGTCGCGCGCGG	1282
RESULT 12			
AE008240/c			
LOCUS	AE008240	14945 bp	DNA linear BCT 18-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 44 of 187 of the complete sequence.		
ACCESSION	AE008240	AE007870	
VERSION	AE008240.1	GI:15158760	
KEYWORDS			
SOURCE	Agrobacterium tumefaciens str. C58 (Cereon).		
ORGANISM	Agrobacterium tumefaciens str. C58 (Cereon)		
	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
REFERENCE	1 (bases 1 to 14945)		
AUTHORS	Hinkle, G., Slater, S.C. and Goodner, B.		
TITLE	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall		

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

Disease in Plants
Unpublished
2 (bases 1 to 14945)

Hinkle, G., Slater, S.C. and Goodner, B.

Direct Submission

Submitted (14-AUG-2001)

Bioinformatics, Cereon Genomics, 45 Sidney

Street, Cambridge, MA 02139, USA

Approximately 800 bp of telomeric sequence missing from the left

end of the chromosome and 200 bp missing from the right end.

Location/Qualifiers

1..14945

/organism="Agrobacterium tumefaciens str. C58 (Cereon)"

/strain="C58"

/db_xref="taxon:181661"

98..865

/gene="AGR_L_810"

98..865

/gene="AGR_L_810"

/note="HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN GLPC-AIS

INTERGENIC REGION"

/codon_start=1

/transl_table=11

/product="AGR_L_810p"

/protein_id="AAK88977.1"

/db_xref="GI:15158761"

/translation="MTQDDSERYPALDKGLDLELARTDGGLTQIEIAKAIGKSP

NELRLDLVRGGYVQRDGDGRFSLTKMGLAHFHAPIRRLVFAAPVMDRFSASA

EQACHLAVYDRGSVVVIAQESATYAMSMRVGAQMSLFTGSGHLLAFQTEAOREI

MTTEQVRGSGETVPADLAERLRQVRSNGFESMDLSLOTAGVRNISAPILTLDGTALAVI

TCPYITLGGTAPTRETCEQLIRDAAKRISEVVTGNV"

complement(972..1937)

/gene="AGR_L_812"

complement(972..1937)

/gene="AGR_L_812"

/note="transaldolase PA2796 {imported} - Pseudomonas

aeruginosa (strain PA01)"

/codon_start=1

/transl_table=11

/product="AGR_L_812p"

/protein_id="AAK88978.1"

/db_xref="GI:15158762"

/translation="MTSKLEQLRAITVVADTGDIENAVAREKPVDCNTPTVLKALG

TPAFADAKEAVAWRGKGGQSDAVVAADVADRLAISVGAALAGLVPGRVSTVEDADLS

FDQTASITKARHIIASIKERGIERERILIKLASTWEGIKAAEVLOSIGIDCNLTLLFS

QAQATACAEAKAFILSPFVRILDWYKKSTGETYTAETDPGVVSRYSIYKANGIG

TVMGASFRNVEIGIEALACDRLTISPALLELDKDKGLERKLSPDNVKAEALQSLD

EKAERWMNEDAMATEKLSGIRLFAKDLVTLRENVRELTAATAA"

2133..3164

/gene="AGR_L_814"

2133..3164

/gene="AGR_L_814"

/note="(AP001509) transcriptional regulator"

/codon_start=1

/transl_table=11

/product="AGR_L_814p"

/protein_id="AAK88979.1"

/db_xref="GI:15158763"

/translation="NAKTVAKLRRETHHTYSEASLRLRAAMLYYNOGMTOKDVAERL

GISRSTVIRLDEAMKRSVQIWIWEGIEDFVSLAGOLEAAVGLDEAVTIIPSPGPSA

GPSTASPSTAKTSAGCTAKAVGLAGLQGLFSEVPNGATIGVGCWRTMTASLSRPP

RRNCKVSVSLGGIVAVQPLDPTWRLASALGAECYNFLAPLLVDSVETKRALIEK

CGLATYDLAETDLAIYSCGDIPGHSTLSGEGFSRETRELVDAGVCVDFMFMFID

AEGRSVDPHPINORAMAIDTLRKAKHIVLASGGAAHRAIATIRATIKRICNTLIIDEA

AAALMELV"

complement(3161..3448)

/gene="AGR_L_815"

complement(3161..3448)

/gene="AGR_L_815"

/note="HYPOTHETICAL 11.0 K0 PROTEIN Y4Kp"

/codon_start=1

/transl_table=11

/product="AGR_L_815p"

/protein_id="AAK88980.1"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

```
/db_xref="GI:15158764"
/translation="MRLIWRVRLRELDAGVYTAERNPRAAAKVVRAIHOTTARLLS
DNPLGRSGIEGTRRLVPLGPIVIVRYTEEQVDIVFVQHAAREWDDA"
complement(3454..3690)
/gene="AGR_L_816"
complement(3454..3690)
/gene="AGR_L_816"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_816p"
/protein_id="AAK88981.1"
/db_xref="GI:15158765"
/translation="MKSTIELSDDIKRRLLDIAERSNSTPSRIEDALSHGRLAWOE
KWTSGVRAGLAEDAGEFVTEEDVNLKYAKA"
complement(3743..4615)
/gene="AGR_L_818"
complement(3743..4615)
/gene="AGR_L_818"
/notes="O6-methylguanine-DNA methyltransferase PA2118
(imported) - pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_818p"
/protein_id="AAK88982.1"
/db_xref="GI:15158766"
/translation="MNANIMNEDITPIGSDYDVRGVIELTLDYREOPSLEAIAAR
LGSPTLOKTFTRWAGLSKAPLOAVTLQHAKEELLEEDPLLETIEVMSGPRL
LDFVTHEAMSPENKAKGGGLIIRYGFHSPGLALVNIITDRGLACFADRGDERA
CFEDMAGRNADYDREATAFYAARIFEPAMWTADKPLRVLLGLDFQVRWKSLL
KIPMGRAVYTSNACDIOGPTASRAVGAAGVANSFVVFCHRAVKGSGALTGYHWWL
TRKRMGLGWETGKA"
complement(4882..5364)
/gene="AGR_L_819"
complement(4882..5364)
/gene="AGR_L_819"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_819p"
/protein_id="AAK88983.1"
/db_xref="GI:15158767"
/translation="MDTINDOPIFAEELVPHRSIGRRGERLLLLTGFACLIYGGFEL
ATGAMPILGFGLDGLLLYIAFRANYRAAKAEVSVSRNLSIRKFSAPGRMVEYRF
NPWFARVRVRHDEIGIVSMHVTGEGRATIDIGSLPNDPDRSFAPKAFGALATVRRRM
"
5402..6184
/gene="AGR_L_820"
5402..6184
/gene="AGR_L_820"
/notes="endonuclease III PA3495 (imported) - pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_820p"
/protein_id="AAK88984.1"
/db_xref="GI:15158768"
/translation="MTVAKKRSSQTLLKKSIPAORRKPVRVKTAYSKDELTEIFRRS
IQREPGELEHTNPTLLVVALSAQATDVGNRATRALEFKVADTPEKMLALGEEQL
IGHKTLGLYRNKRNKIALSQMLIDNFGVEVPKTRBELVTLPGVGRKRTANVMSAF
GVPLTAVTHVFRANRLCLAPKGTITDEVEDLVRIIPEQYLFHAHHLILHGRYCK
ARKPECRCVIADIKSEKPTFDIPAPLVELPAOLFGAAGE"
complement(6146..6883)
/gene="AGR_L_822"
complement(6146..6883)
/gene="AGR_L_822"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_822p"
/protein_id="AAK88985.1"
/db_xref="GI:15158769"
/translation="MIFDAARLAFGNLFAAETRSVFWKVLGLTLLVLAALWFAIRSF
```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```
IYFALPWIDTLIPGVPDWAGWLSFVFAIFAGIGIALALALISPTAVIAGLFLDDVA
DVYKONYPDDPGKAMPIGEAVSSIKFFCVIAGNLVALLLLVPGVNLIAFELVN
GYLLGREFEFAAMRRFSPADARLFRVYKHRTVTVMAGLVIAAFLAVFLNLTLPLEAA
SLMVLHKAYSRDPSPFAAGTEQLRG"
7179..8012
/gene="AGR_L_824"
7179..8012
/gene="AGR_L_824"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_824p"
/protein_id="AAK88986.1"
/db_xref="GI:15158770"
/translation="MVBKLLNLTRAMKAGTPSPERKIYKYLIEHLDELPPETAOTLAA
KLSLSPMTVGRFLSLCYROFSDIRADLRHAEVTVGADOPVTAEGGHRHSPFSQLLS
QQIQAIQTAFDMTQTWRLAMNEITSADDOVCLATSPGLDAGRHFGYRLLCEGRGHIH
YLGSDNAYVALWDYDPARTLLVIMDCGNLLPLORLSTARKGGRYRLLLTIRFYEW
GPESVDLCAMPQSQNGGQGLQLVLSLEFTLCALSAGADGAKRVKNLAALKRSLR
S"
complement(8117..9694)
/gene="AGR_L_826"
complement(8117..9694)
/gene="AGR_L_826"
/notes="D-ribulokinase (EC 2.7.1.47) - Klebsiella
pneumoniae"
/codon_start=1
/transl_table=11
/product="AGR_L_826p"
/protein_id="AAK88987.1"
/db_xref="GI:15158771"
/translation="MMRONLVADVGTASARAGIFDPAGRLRLARSIHPLMQRPRENH
AEHSDTIWNACVIAKAAALADAGVLQPSIAAIGFATCSLVIRDERGEPVSVSTDD
DREDTIWLQHRAIGEDRLTASGRVLDFAGNSVSPQMOKMLWLKTHMPOQSMRM
SPAFDLADELTWKATGSNQRNCTQTAKWNLAEHFGWQADYLAFLGDLDLKERAGL
PETVMEFGAIGPLUSAAEELGDTGCQVAAGMIDATAGALGALGGCLAEDEGKHA
LIAGTSSCLVAMSTQPMPSLWGPVQWAVLPQHWLVVEGGSATGALLDHIVRMHAAG
GEPDTALHARIVARVTELRELEGFADRLHLPDFHGNRSPLADPHAVGVVSGLTLD
TSFDSLRLRYWTAIALGARHVLDMAREFGYAVETLHTGTGHVKNPLLMELYADV
GKRVIVPATADAVLGLTAMTAAATAGGVHASLAAAGAAAMPYPCNAETISGNPALAAHYERD
```

Query Match 6.1%; Score 112; DB 1; Length 14945;

Best Local Similarity 48.1%; Pred. No. 4.5e-23;

Matches 415; Conservative 0; Mismatches 433; Indels 12; Gaps 3;

Qy 742 GATTTACAAACGCTCATTTGCTGAAGGGGAGATCCGAGACAAAGAGTATGCTTTCACC 801

Db 11629 GATACCCAGGCCATCATCGATAGCTTCCGATCCGAAATCCGATCGCTTTTCGATGACG 11570

Qy 802 GTACCCGAAGGGGGTACTACTGTAATCCAGTCAACAACTCGAAATTAACAATCCTGAT 861

Db 11569 ATTACCGAGGGGGTTATTTTCATCGATGCTCCGCTCCTTCAATCCGAGCATCCGGCC 11510

Qy 862 TTAGCGGCAGATCTTAAAGGGGGATGC-----AAACAAATTTACGGTGTATTACCCGT 915

Db 11509 ATTGCGGAAGCGGCAAGAACCCGCTCCGATCCGAAACCGTGTTCGTTTCATCGTTGCC 11450

Qy 916 ATCCTCGAAGCGCGTATGCAATAAACCGGACCACTAACCCCTGCTGAATTCGGAATAAC 975

Db 11449 GGGCTGAAGGGCGGGCGGATGAAGACCTTCAGCCCTTCACCGTATGCTTCGCGACAAC 11390

Qy 976 GTGCCCAATAATGTGAACGTTTCCATGATGCGCTGGTGTGAGTTTCTCCAGCATTAACGTCG 1035

Db 11389 ATTCCCCACATGGCAAGGTGACGAAACCGGTTGGTGGCGTGGCAGCGCTCTC--C 11333

Qy 1036 AAACAGGATGTCATCGACTGGCTGAGTACAATAACACATTCGCCGAATACCATGTTGATC 1095

Db 11332 GACCCGGGCTTTGCCAACTGGATCGCGAAAAATGCGCTTTCCGAAATTCATGTCGAC 11273

Qy 1096 CGCATTTACGCTCGTCCGGCAGCAGACTTCCGGCAGCGATCAAGCTCAACGGGTATT 1155

Db 11272 CGCATACCCCGCCACCGGCGAGCGTGAACGCAATATCGCCCGTGACGATTCGCGCATC 11213

OY 1156 GCGGATAAAGCCGGTAAATGGCGGAACCTTTATCCAGTGGGTGCTGGGAAGATAATTTC 1215
Db 11212 GAGGATAATGGCCGCTCTTCGCGAAGAGTTCAAGCAATGGGTGATGGAGGATAATTTC 11153
OY 1216 CGTGATGTCGTCGCGCACTGGAGAGGTCGGTGTGCGAACTGGTGGCGTCGGTAAATCCCC 1275
Db 11152 CCGCGAGGCGTCCGCGCTCGAAGAAGCGGGGTCAGTTCTGTCGAAGGATGTTGGCGCC 11093
OY 1276 TATGAAGCGGGAAGATTCGCAATCTTAACTCTTACACAGTTGTCACCTCGCTGGGAGGT 1335
Db 11092 TAGGAGCACATGAAGATCGCATCTCTGAATGTTGGCCACGGCGCAATCGCTATCGGCA 11033
OY 1336 ACCTTAATCGGTCAAAAATATATCCAGGAAGACACAAATGACCGATTTTATCTATCAGATT 1395
Db 11032 GCGCTCTCGACATCTATTTCTGTCATGAGGCGATCGCGTATCGCGGGGTTT 10973
OY 1396 GCGGACCGTACGTGACGGAAGATGTCATTCTTGTCTT---GGCGATAACGATATCGAT 1452
Db 10972 CTGGCCAACTCGMAAAGACGAGATCATCCGGTATTTCGCGCGGTACCTGACACCGAC 10913
OY 1453 TTGCCAACCTACCGTGATGTTGTACTCAAGCGTTTTTACCAATCCACATATTCAGGACACC 1512
Db 10912 CTCAACGACTATTTCGGGCTGATCGAGCGCGTTTCTCAACCCGAAGATTGCGGATACC 10853
OY 1513 ACCAACGGTGCCTCGGATGTTTCTCGAAATTCCTCGCGGATGATTCGCCCCACACATG 1572
Db 10852 ATTCCGCGCTCGCGAGGATGTTTCAACCGCGCAAAATTCATCTCGCCCTCACT 10793
OY 1573 CGAGAGTGTACACGAGGCG 1594
Db 10792 TTAGATCGGTGCGCGCGCG 10771

RESULT 13
LOCUS SCE9669
DEFINITION Saccharomyces cerevisiae chromosome V cosmid 9669, 8334, 8199, and
lambda clone 1160.
ACCESSION U18795.1
VERSION U18795.1 GI:603241
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae

REFERENCE
AUTHORS Dietrich,F.S., Mulligan,J., Hennesy,K., Yelton,M.A., Allen,E.,
Araujo,R., Aviles,E., Berno,A., Brennan,T., Carpenter,J., Chen,E.,
Cherry,J.M., Chung,E., Duncan,M., Guzman,E., Hartzell,G.,
Hunnicke-Smith,S., Hyman,R.W., Kayser,A., Komp,C., Lashkari,D.,
Lew.H., Lin,D., Mosedale,D., Nakahara,K., Namath,A., Norgren,R.,
Oefner,P., Oh,C., Petel,F.X., Roberts,D., Sehl,P., Schramm,S.,
Shogren,T., Smith,V., Taylor,P., Wei,Y., Botstein,D. and Davis,R.W.
The nucleotide sequence of Saccharomyces cerevisiae chromosome V
Nature 387 (6632 Suppl), 78-81 (1997)
737313264

TITLE
JOURNAL
MEDLINE
PUBMED 9169868
REFERENCE 2 (bases 1 to 50849)
AUTHORS Dietrich,F.S.
TITLE The sequence of S. cerevisiae cosmids 9669, 8334, 8199, and lambda
clone 1160
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 50849)
AUTHORS Dietrich,F.S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1994) Genetics, SGD, Stanford University, Palo
Alto, CA 94305-5120
REFERENCE 4 (bases 1 to 50849)
AUTHORS Jia,Y.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-1997) Genetics, SGD, Stanford University, Palo
Alto, CA 94305-5120
REFERENCE 5 (bases 1 to 50849)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Jia,Y. and Cherry,J.M.
Direct Submission
Submitted (04-MAR-1997) Department of Genetics, School of Medicine,
Saccharomyces Genome Database, Stanford University, CA 94305-5120,
USA
6 (bases 1 to 50849)
Jia,Y. and Cherry,J.M.
Direct Submission
Submitted (12-JUN-1997) Department of Genetics, School of Medicine,
Saccharomyces Genome Database, Stanford University, CA 94305-5120,
USA

Sequenced by: Stanford DNA Sequence & Technology Center
855 California Avenue
Palo Alto, CA 94304, USA
Curated by: Saccharomyces Genome Database
URL: http://genome-www.stanford.edu/
e-mail: yeast-curator@genome.stanford.edu

Neighboring Sequencing:
The 3' end of this sequence overlaps with GenBank Accession Number
U18779.

FEATURES
source

source

source

source

source

gene

CDS

gene

CDS

/note="similar to hypothetical 65.4 kD protein in
subtelomeric y' repeat from S. cerevisiae, Swissprot
Accession Number P24088"
/codon_start=1
/evidence=not_experimental
/product="Yel076cp"
/protein_id="AAB65011.1"
/db_xref="GI:603242"
/translation="MQASLPGEKKVDTERLKRDLCPKRPKPIETKYSQICNDMMNKKDR
LGDILHILRACALNFGAGPRGGAGDEDSITNEEPIPSVDEHGLKVAQETLKDSYR
RLRLKTLDAVKALLVSSCACTARDLDIFDNNNGVAMKWKILYHEVAQETTLKDSYR
ITLVPSSDGIPTLTVIOFSYSLPLVLSATYTSMIQQDASNCITLITTRVHRSLD"
/complement(1987..2637)
/gene="YEL076C"
/complement(1987..2637)
/gene="YEL076C"
/note="similar to hypothetical 65.4 kD protein in
subtelomeric y' repeat from S. cerevisiae, Swissprot
Accession Number P24088"
/codon_start=1
/evidence=not_experimental
/product="Yel076cp"
/protein_id="AAB65011.1"
/db_xref="GI:603242"
/translation="MQASLPGEKKVDTERLKRDLCPKRPKPIETKYSQICNDMMNKKDR


```

IGDLHLILRACALNFGAGPRGGAGDEEDRSITNEEPIIPSVDEHGLKVKLRSPNTP
RLRLKTLDAVKALVSGCCTARDLIDFDDNNGVAMMKWIKILYHEVAQETTLKDSYR
ITLVPSSDGIISVCGKLGELNCFHPASPFCVIEKFCSENSELTSLWKTILFEK"
2393..3004
/gene="YEL076W-C"
CDS
2393..3004
/gene="YEL076W-C"
/codon_start=1
/evidence=not_experimental
/product="Yel076w-cp"
/protein_id="AAB65035.1"
/db_xref="GI:1857035"
/translation="MGSSSFVIDRSSSPAPRGPAPKLSAHARKIICKISPNRSFLFI
ISLHICKEYFMSGLRHSRFSRVSSTFFSPGLACIAHLRVGQIVPIPPYGAFLK
TPYNRCAGKYSSESTHRRVVRSPTRVFTFQDETOLIIIVSSVEVKRKGIVILSI
EFQSMHLKORVDHVVCLGLVRPGLNVLRLQQTSGTFLHFSRP"
complement(2868..3236)
/gene="YEL075C"
CDS
complement(2868..3236)
/gene="YEL075C"
/translation="similar to hypothetical 65.4 kD protein in
subtelomeric Y' repeat from S. cerevisiae, Swissprot
Accession Number P24088"
/codon_start=1
/evidence=not_experimental
/product="Yel075cp"
/protein_id="AAB65012.1"
/db_xref="GI:603243"
/translation="MKVSDRKEFEKANDEFESALNNKNDLVHCPSTILFSPISITEVR
SFYEDRSGLKLVKFRGAMDRKRSFEKVISVWGNVKKFLTFVEDEPDFGGPI
PSNKRDRGLHVSSAYFEIQ"
3649..3987
/gene="YEL074W"
CDS
3649..3987
/gene="YEL074W"
/codon_start=1
/evidence=not_experimental
/product="Yel074wp"
/protein_id="AAB65013.1"
/db_xref="GI:603244"
/translation="MDLRSALKETNIKSKYKHLFLIRVTTSRFEVKEPFLAYLHMARQ
KDKLLPKCITPPNKKADPFTNDVSSYFTHTHPCSPSLVYTYVNTKEKSPSKPKH
KNILPFNETK"
complement(4753..5076)
/gene="YEL073C"
CDS
complement(4753..5076)
/gene="YEL073C"
/codon_start=1
/evidence=not_experimental
/product="Yel073cp"
/protein_id="AAB65014.1"
/db_xref="GI:603245"
/translation="MVNLNVLTNATATLSAMSNVTVPLETYFHFDEASGFDYVLNV
SVIWMNETLYETRIVPALINRVRELDHMEANDPSPVTNPYETSGYAFSVVPLMG
NMKVA"
11243..11938
/gene="YEL072W"
CDS
11243..11938
/gene="YEL072W"
/codon_start=1
/evidence=not_experimental
/product="Yel072wp"
/protein_id="AAB65015.1"
/db_xref="GI:603246"
/translation="MSACPNIIVLPVELLNSSKDTKYSLYTTINRGYDVPRLKYGI
IVSPRVHSLFTLSDLGFDKNIEKSSLYLLNDPLAYPNEHEHFEQURKGTNDLSL
PTYIPKQVQFTEAFDESHTLATIGYKPNKSEYEITGFTSMGNGYKLFNFYSVIHM
NRSHKCRKRVADIIMEHDLGLGYEKKLGFVEVQRKVLKEHQKVFDDKVDFTKDFH
VIMKIKELGNHRL"
13878..15368
/gene="YEL071W"
CDS
13878..15368
/gene="YEL071W"

```

```

/note="similar to D-lactate dehydrogenase from S.
cerevisiae, Swissprot Accession Number P32891"
/codon_start=1
/evidence=not_experimental
/product="Yel071wp"
/protein_id="AAB65016.1"
/db_xref="GI:603247"
/translation="MTAAHPVAQLTAETAYPKVRNPFKVLDSLAYFRSILSNDIEI
LNSGAPPELASFNODMKYRGQSNLILPNSDTKVKIMKYCNCKKLAVVPOGGNTD
LVGASVPVFEIVLSLRNMKNVROFDPVSGTFCDCAGVVMRDAHGLFHDHDIPLDL
PSRNCQGVGVSTNAGLNFELRYGLSHGNVLGLEVLNPGELIISINAKLRDNTGYD
LKQFTIGAEGTIGVTVYVIAAKAPKALNAVFFGIENFTVQKLFVAKSELSEILS
AFEMDRSIECTIETYLKDLPPFLPQHNHNFYLIETSGSNRKHDDDEKUTAFKLTDTDS
KLISEGMAKADKADRLWTKRVSPTACNSGYKMYKIDMSLQDLXLYSVSAVTERL
NAAGLIGDAPKPVVKSCYGHVGDGNIHLNAVREFTKQIEDLLPFPVVEYIASKKGS
ISABHGIGFHKKGLHYTRDIEIRFMKDINKHDPNGLNLPYKI"
Query Match 6.0%; Score 111.4; DB 8; Length 50849;
Best Local Similarity 49.6%; Pred. No. 8.8e-23;
Matches 314; Conservative 0; Mismatches 316; Indels 3; Gaps 1;
QY 949 CCACCTAACCCCTGCTGAATTGCGATACGTCGCGCCATAATGGTGAACGTTTCCATGATGCG 1008
Db 17682 CCATTACATATTATGTCATGTGACAACTATGCCCAAAATGGTGTACAGTAAGACCATG 17741
QY 1009 CTGGTTGAGTTTCTCCAGCTAACTGGCAACAGGATGTCATCGACTGGCTGAGTACAAAT 1068
Db 17742 CTTGTTGCATTTGCCCAAGTTAAAGAGGATGAGAAATTCGCCGCTGGATTGAAGACAAG 17801
QY 1069 ACCACTTGGCCGAATACCATGTTGACCGCATTCACCTCTGTCGCGCAGCAGAACTTCG 1128
Db 17802 GTTACTTCTCTAACAGCATGTTGGACCGTGTGACCCACCTGTGTACCGATAAGACGCT 17861
QY 1129 GCACGGATCAAGGCTCAAAACGGGTATTGCCGATAAAGCCGCGTAATGGCGGAAACCTTT 1188
Db 17862 AAATACGTTGCTGCACACTGGGGAATCAAGATCAATGTCCCGTTGTCGACGAACTTTC 17921
QY 1189 ATCCAGTGGTCTGGAAGATAAATTCCTGTGATGTCCTCGGCACTGGAGAGGTCGGT 1248
Db 17922 ATCCAATGGTCTTGAAGACAACTTCTCCGATGGCCGTCCTCCACTGGGAACCTTGTGGT 17981
QY 1249 GTCGAACCTGGTGGCGTCGTAATCCCTATAGAGAGCGCAAGATTCGCATTCTTAATCT 1308
Db 17982 GTTCAGGTCGTCGAAGGATGTCGATTCCTACGAATTTGAAGTTGCGTCTACTTAACGGT 18041
QY 1309 TCACACAGTTGATCGCTCGCCGCGGAGGTATTAATCGGTCAAAAATATATATCCACGAAAGC 1368
Db 18042 GGACATTCGCTATGGGATATTGGGATATTGGGAGGCTACACCTATATACATGAGTT 18101
QY 1369 ACAATGACCGATTTTATCTATCATGATTGCCGACCGCTAGTCAGCGAAGATGTCATTCT 1428
Db 18102 GTCACAGCCCCAACTATCAACAAGTATATCCGTTGTTGATGCGTCGAGGAAGTTATCCCA 18161
QY 1429 TCGTTG---GGCGTAACCGGTATCGATTGGCCAACTACCGTGATGTTGTACTCAACGCT 1485
Db 18162 TTATTCCTTAAAGTGCAGGCTGTGATTTCGAAGAGTACACTGCATCAGCTGTTGGAAAGA 18221
QY 1486 TTTACCAATCCACATATTCAGGACACCAACCAACGCGTCGCTCGGATGTTCTCTCGAA 1545
Db 18222 TTCTCCAATCCAGCAATTCAGGACACTGTGCGACGATTTGTTGTTGATGGCTCTGTTAAG 18281
QY 1546 ATTCGGCGGATGATTGCCGCCACACATTCGCGAGAG 1578
Db 18282 ATGCCTAAGTATGTTTGGCCATCAATTTACGAG 18314
LOCUS AE005449 10449 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 68
of 290.
ACCESSION AE005449 AE005174
VERSION AE005449.1 GI:12516490

```

KEYWORDS	Escherichia coli O157:H7 EDL933.	gene	complement(2369. .4060)
SOURCE	Escherichia coli O157:H7 EDL933		/gene="frua"
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	CDS	/note="23425"
REFERENCE	1 (bases 1 to 10449)		/complement(2369. .4060)
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		/gene="frua"
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7		/function="regulator; Degradation of small molecules:
JOURNAL	Nature 409 (6819), 529-533 (2001)		Carbon compounds"
MEDLINE	21074935		/note="Residues 1 to 563 of 563 are 99.46 pct identical to
PUBMED	11206551		residues 1 to 563 of 563 from Escherichia coli K-12 Strain
REFERENCE	2 (bases 1 to 10449)		MG1655: B2167"
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		/codon_start=1
TITLE	Direct Submission		/transl_table=1
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		/product="PTS system, fructose-specific transport protein"
FEATURES	Location/Qualifiers		/protein_id="AAG57305.1"
source	1. .10449		/db_xref="GI:12516493"
	/organism="Escherichia coli O157:H7 EDL933"		/translation="MKTLLIIDNLGQARAYMAKTLTGAARAKKLEIINDPNDAEMA
	/strain="EDL933"		IVLGDSPNDSALNGKNVLDGDISRAVHPELFLSEAKGHAKPYTAPVAATAVAASG
	/serotype="O157:H7"		PKRVVACTPGTGAHFTFMAEALETEAKKRGKWKVKVETRGSVGAGNATIPREVAAD
	/db_xref="taxon:155864"		LVIYAADIEVLAKFACKPGRYSTGLAKLKTQELDKVAEATPVPEAGKATATSE
gene	/note="enterohaemorrhagic"		GKNESAGYRHLLTGVSYMLPMVAVAGGLCTALSFAFGIEAFKEPFTLAALMQIGGS
	complement(78. .1016)		AFALMVPVLMGAIATPSIADRPGLTGLIGGLAVLSTGSGFIIAGFLAGYATKLAS
	/gene="yeiN"		TOLKLPQSMALKPILIPILDISLVGLAMLYLIGKPVAGILGLELTHWQTMGTANV
	complement(78. .1016)		LLGAILGGMCTDMGPGVNAAYAFGVGLLSTQTYGPMAAIMAAVGPPLAMGLATMV
	/note="23422"		ARRKFDQAQQEGKAALVLGLCFISEGAIPFAARDPMRVLPCPCIVGAGATGAISMAIG
	/gene="yeiN"		AKLMAHGGLFVLLIPGATTPVLGYLVAILAGTLVAGLAVAFLEKRPVEDAVAKAA"
	complement(78. .1016)		complement(4077. .5015)
	/gene="yeiN"		/gene="fruk"
	complement(78. .1016)		/note="23426"
CDS	/function="orf; Unknown function"		complement(4077. .5015)
	/note="Residues 1 to 312 of 312 are 99.35 pct identical to		/gene="fruk"
	residues 1 to 312 of 312 from Escherichia coli K-12 Strain		/function="enzyme; Energy metabolism, carbon: Glycolysis"
	MG1655: B2165"		/note="Residues 1 to 312 of 312 are 100.00 pct identical
	/codon_start=1		to residues 1 to 312 of 312 from Escherichia coli K-12
	/transl_table=11		Strain MG1655: B2168"
	/product="orf, hypothetical protein"		/codon_start=1
	/protein_id="AAG57303.1"		/product="fructose-1-phosphate kinase"
	/db_xref="GI:12516491"		/protein_id="AAG57306.1"
	/translation="MSELKISPELLQISPEVQDALNKKPVVALESTIISHGMPFPQN		/db_xref="GI:12516494"
	ACTALEVETTRKQGVNATTAIGGVKWKVGLSKEEIELLRGHNVTKYVRDLDFV		/translation="MSRRVATITLNPAYDLVGPCEPETERGEVNLVKTGLHAAGKGIN
	VAAKRGATVASTMTAALAGIKVFATGGIGGVHVRGAEFTDISADQLANVTV		VAKVLKOLGIDVTYGGELGKNDQDFQOLESELGIANRFOVQGRPRINVKLTERKQGE
	VCAKASILDGLTETXLETGVPGLIYQTKALPAFFCTSPFDVSIRLDSASEIARA		VTDFNFSGFVTPADWERFVDTLSLUGQDMVCVSGSLPSGVSPEAFDTWMTLRKSG
	MAVKWQSLGGLVYVNPDPQFAMPETRNAVIDQAVAEAAEQGVIGKESTPTFLAR		CPCIIIFDSSREALVAGLKAAPLWVKPNRBEIHWAGKRLPEMKDVIIEAALHREQQIA
	VAELTGGDSLKSNLQLVFNNAIILASEIAKEVQLAG"		HVVISLGAEGALVYNAWGEWIAKPPSYDVVTVTGAGDSVMVGLIYGLLMRESSEHTLR
	complement(1004. .1945)		LATAVAALAGVSNVGTIDRPLQLAAMMARVDLPFN"
	/gene="yeiC"		complement(5015. .6145)
	/note="23423"		/gene="frub"
	complement(1004. .1945)		/note="23427"
	/gene="yeiC"		complement(5015. .6145)
	/function="putative enzyme; Not classified"		/gene="frub"
	/note="Residues 1 to 313 of 313 are 99.04 pct identical to		/function="enzyme; Transport of small molecules:
	residues 1 to 313 of 313 from Escherichia coli K-12 Strain		Carbohydrates, organic acids, alcohols"
	MG1655: B2166"		/note="Residues 1 to 376 of 376 are 99.73 pct identical to
	/codon_start=1		residues 1 to 376 of 376 from Escherichia coli K-12 Strain
	/transl_table=11		MG1655: B2169"
	/product="putative kinase"		/codon_start=1
	/protein_id="AAG57304.1"		/transl_table=1
	/db_xref="GI:12516492"		/product="PTS system, fructose-specific IIA/fpr component"
	/translation="MREKDYVVIIGSANIDVAGYSHESLNAYDSNPCKTKFTPGGVGR		/protein_id="AAG57307.1"
	NIAQNALLGNKALLSNVGSDFYQSLITQTQSGVYVDKCLIVRGENTSSYLSLLD		/db_xref="GI:12516495"
	NIFEMLVAINDMNISITAYLAQHRFEIQRKAVITVADNCISEEALAWFLDRAANVP		/translation="MFOLSVQDTHPGEKAGDKKEATRQVAAALVQAGNVAEGVNGCML
	VFVDPVSAWKCVKVRERLSQTHLKNRLEAEFTLSGLSREDVAKVAWHPQGNL		AREQOTSTFLNGITAIPIHGTITDRDQVLKTVGVQVFQPEGVTWGDCQVAYVAIGIAAS
	RLVLSMGDGGYVYIDISGWSGAPIKNTINVTGAGDMAAGLASCWVDGMPFAESV		SDEHLGLLRQLTHVLSDDSAVEQLSATTAEARLLMLKEQSEQLKDNELMLTDIIV
	RFAQCSMSALSCYTNPNPLDSIANVLSVNEBCLN"		ASDLTLTQALNAARLKEAGVADATFVTKAINEQLPNLGGQISLWLSAEGLNLSATIAV
			RAANAFVDGETAAFLVSVAMNDQPTIAVLKRLDLLNDKADRLKADAATLLALLT
			SDAPATDDVISAEEFVRNEHGLHARPGTMLVNTIKOFNSDITVTYTLNLDGTCKPANGRSL
			MKVVALGVKKHRLRFTQAQADAEQALKAIGDAIAAGLGBGA"
			6513. .7694
			/gene="yeiO"
			/note="23428"
			6513. .7694
			/gene="yeiO"
			/function="orf; Not classified"

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:12:47 ; Search time 409.996 Seconds
(without alignments)
10150.564 Million cell updates/sec

Title: us-09-802-208b-1

Perfect score: 1848

Sequence: 1 atgaacgaacattacatg.....ttacgcgttaataactaa 1848

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1848	100.0	1848	22	AAD16810 Escherichia coli s
2	105	5.7	4115	21	AAAS7430 Gluconobacter oxyd
3	101.8	5.5	1458	20	AAZ35672 Gluconobacter subo
4	87.6	4.7	9810	20	AAZ32025 Human METHI relate
5	87.6	4.7	9810	22	AAC90082 AF018073 cDNA clon
6	80.6	4.4	921	23	AAAS92759 DNA encoding novel
7	73.4	4.0	2810	23	AAAS78056 DNA encoding novel
8	68.8	3.7	1509	22	AAH65091 C glutamicum codin
9	68.8	3.7	1632	22	AAF71527 Corynebacterium gl

c	10	68.8	3.7	349980	22	AAH64966 C glutamicum codin
c	11	66.8	3.6	1182	23	AAAS93068 DNA encoding novel
c	12	61.2	3.3	684	21	AAF12753 Aspergillus oryzae
c	13	43.4	2.3	2193	23	AAAS82343 DNA encoding novel
c	14	42.4	2.3	14041	22	AAH48024 Internal control B
c	15	38.4	2.1	2238	23	AAAS92760 DNA encoding novel
c	16	33.2	1.8	278	22	AAAL01054 Human reproductive
c	17	33.2	1.8	278	23	ABL96521 Human testicular a
c	18	33.2	1.8	4787	20	AAZ24631 Human lung tumor a
c	19	33.2	1.8	4797	21	AAAC65870 Human lung cancer-
c	20	33.2	1.8	4797	24	ABN97444 Gene #3942 used to
c	21	33.2	1.8	4797	24	ABL68585 Kidney cancer rela
c	22	33.2	1.8	4797	24	ABL49089 Human lung tumour
c	23	32.8	1.8	615	23	AAAS1468 Pseudomonas aerugi
c	24	32.4	1.8	2230	24	ABL56160 Maize acyl CoA oxi
c	25	32.4	1.8	2436	23	AAAS4141 Pseudomonas aerugi
c	26	32.4	1.8	3098	23	ABL05639 Drosophila melanog
c	27	32.4	1.8	3271	24	AAI67053 D. melanogaster PA
c	28	32.4	1.8	5166	23	ABL05638 Drosophila melanog
c	29	32.4	1.8	134141	24	ABN83487 Escherichia coli O
c	30	32	1.7	512	24	ABN16299 Human ORFX polyuuc
c	31	32	1.7	2139	23	AAAS6324 Salmonella typhi D
c	32	32	1.7	3725	21	AAZ99500 DNA encoding a mai
c	33	32	1.7	3725	21	AAZ99506 DNA encoding a mai
c	34	32	1.7	3725	21	AAZ99521 DNA encoding a mai
c	35	31.8	1.7	2022	23	ABL09849 Drosophila melanog
c	36	31.8	1.7	2061	21	AACT76801 Human ORFX ORF2356
c	37	31.8	1.7	2083	22	ABA09036 Human PTD014 homol
c	38	31.8	1.7	2084	22	AAI06020 Human reproductive
c	39	31.8	1.7	2084	22	AAI06021 Human reproductive
c	40	31.8	1.7	2084	23	ABL98585 Human testicular a
c	41	31.8	1.7	2084	23	ABL98586 Human testicular a
c	42	31.8	1.7	2977	23	ABL17060 Drosophila melanog
c	43	31.8	1.7	4022	23	ABL09848 Drosophila melanog
c	44	31.8	1.7	23620	23	ABL29054 Drosophila melanog
c	45	31.8	1.7	23620	23	ABL19102 Drosophila melanog

ALIGNMENTS

RESULT 1

AAD16810

ID AAD16810 standard; DNA; 1848 BP.

XX

AC AAD16810;

XX

DT 29-NOV-2001 (first entry)

XX

DE Escherichia coli strain C arabitol dehydrogenase gene.

XX

KW Positive selection system; metabolise; arabitol; ribitol; mannitol;

KW transgenic cell; marker gene; arabitol dehydrogenase; ds.

XX

OS Escherichia coli C.

XX

PN WO200166779-A2.

XX

PD 13-SEP-2001.

XX

PF 08-MAR-2001; 2001WO-US07474.

XX

PR 08-MAR-2000; 2000US-0188291.

XX

PR 15-AUG-2000; 2000US-025595.

XX

(UYGE-) UNIV GEORGIA RES FOUND INC.

XX

PI Parrott W, Lafayette P, Kane P;

XX

DR WPI; 2001-565596/63.

XX

PT Positively selecting transformed cells comprising selectable marker gene and desired gene, from a cell population by using marker compounds

PT	e.g., arabinol, ribitol which confer selective advantage on transformed cells	
PT	XX	
PS	Claim 1: Page 34-35; 37pp; English.	
XX		
CC	The present invention relates to a positive selection system that involves conferring to transformed cells the ability to metabolize arabinol, ribitol and/or mannitol. The positive selection method is used in positively selecting transgenic cells from a population of cells using the positive selection method, the presence of the gene of interest in the genetically transformed cells may be determined without the disadvantages associated with traditional negative selection systems. Positive selection of the transformed cells is achieved without directly damaging the neighbouring non-transformed cells. The transformed cells may be identified by simple visual means without the use of a separate assay to determine the presence of a marker gene. This technique also avoids the release of antibiotics or other dangerous genes into the environment. The present sequence is Escherichia coli strain C arabinol dehydrogenase gene.	
CC	SQ	Sequence 1848 BP; 494 A; 443 C; 473 G; 438 T; 0 other;
XX		
Query Match 100.0%; Score 1848; DB 22; Length 1848;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1848; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	ATGAAGCAAAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG 60
DB	1	ATGAACGAACAATTTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAG 60
QY	61	GGGTGATCTACACCGTTTGACGGTGTAGGGGATFAAACGGTGGAGCATGTCGGGGC 120
DB	61	GGGTGATCTACACCGTTTGACGGTGTAGGGGATFAAACGGTGGAGCATGTCGGGGC 120
QY	121	AATATTCTGAATGATCTGAACATGCTACAGGCACTGATGCCACAGAAAGTCCCTAT 180
DB	121	AATATTCTGAATGATCTGAACATGCTGACAGGCACTGATGCCACAGAAAGTCCCTAT 180
QY	181	GTGCTGAAACCGTCAGCCCGGAAGGGTAAAGCAATATGAAGAGATCACTCAATTCAAG 240
DB	181	GTGCTGAAACCGTCAGCCCGGAAGGGTAAAGCAATATGAAGAGATCACTCAATTCAAG 240
QY	241	AGTTGATACCGTGGCAGCGAGATTTACAAACCGCTGATTTGTAAGGGGCGAGATCCGAAG 300
DB	241	AGTTGATACCGTGGCAGCGAGATTTACAAACCGCTGATTTGTAAGGGGCGAGATCCGAAG 300
QY	301	ACAAAAGTGATGCTTTACCGCTACCGAAGGGGGTACTACTGTAATACCAAGTCAACAA 360
DB	301	ACAAAAGTGATGCTTTACCGCTACCGAAGGGGGTACTACTGTAATACCAAGTCAACAA 360
QY	361	CTGGAAGTTAAACAATCCTGATTAGCGGCAGATCTTTAAAGGGGGATGCAAAACAATTTAC 420
DB	361	CTGGAAGTTAAACAATCCTGATTAGCGGCAGATCTTTAAAGGGGGATGCAAAACAATTTAC 420
QY	421	GGTGTATTACCGGTATCTCGAAGCGGTATGGCAAAATACCGCGGACCATTAACCCCTG 480
DB	421	GGTGTATTACCGGTATCTCGAAGCGGTATGGCAAAATACCGCGGACCATTAACCCCTG 480
QY	481	ATGAACGAACAATTTACATGGCTGCACATCGGTTAGGTTCTTTTCATCGCGCACATCAG 540
DB	481	ATGAACGAACAATTTACATGGCTGCACATCGGTTAGGTTCTTTTCATCGCGCACATCAG 540
QY	541	GGGTGATCTACACCGTTTGACGGTGTAGGGGATFAAACGGTGGAGCATGTCGGGGC 600
DB	541	GGGTGATCTACACCGTTTGACGGTGTAGGGGATFAAACGGTGGAGCATGTCGGGGC 600
QY	601	AATATTCTGAATGATCTGAACATGCTGACAGGCACTGATGCCACAGAAAGTCCCTAT 660
DB	601	AATATTCTGAATGATCTGAACATGCTGACAGGCACTGATGCCACAGAAAGTCCCTAT 660
QY	661	GTGCTGAAACCGTCAGCCCGGAAGGGTAAAGCAATATGAAGAGATCACTCAATTCAAG 720
DB	661	GTGCTGAAACCGTCAGCCCGGAAGGGTAAAGCAATATGAAGAGATCACTCAATTCAAG 720

QY	721	AAGTTGATACCGTGGCAGGCGAGATTTACAACCGCTGATTGCTGAAGGGGCGAGATCCGAAG 780
DB	721	AAGTTGATACCGTGGCAGGCGAGATTTACAACCGCTGATTGCTGAAGGGGCGAGATCCGAAG 780
QY	781	ACAAAAGTGATGCTTTACCGCTACCGAAGGGGGTACTACCTGAATACCACTCACAA 840
DB	781	ACAAAAGTGATGCTTTACCGCTACCGAAGGGGGTACTACCTGAATACCACTCACAA 840
QY	841	CTGGAAGTTAAACAATCCTGATTAGCGGCAGATCTTTAAAGGGGGATGCAAAACAATTTAC 900
DB	841	CTGGAAGTTAAACAATCCTGATTAGCGGCAGATCTTTAAAGGGGGATGCAAAACAATTTAC 900
QY	901	GGTGTATTATACCGGTATCTCGAAGCGGTATGGCAAAATAAACCGCGACACTTAACCCCTG 960
DB	901	GGTGTATTATACCGGTATCTCGAAGCGGTATGGCAAAATAAACCGCGACACTTAACCCCTG 960
QY	961	CTGAATTTGGGATTAACGTGGCCCATTAATGGTGAACGTTTCCATGATGGCCTGGTTGAGTTT 1020
DB	961	CTGAATTTGGGATTAACGTGGCCCATTAATGGTGAACGTTTCCATGATGGCCTGGTTGAGTTT 1020
QY	1021	CTCCAGCTAACTGGCAACAGGATGTCATCGACTGCTGAGTACAAATACCACTTCCCGC 1080
DB	1021	CTCCAGCTAACTGGCAACAGGATGTCATCGACTGCTGAGTACAAATACCACTTCCCGC 1080
QY	1081	AATACCATGTTGACCGCATTTACGCCCTCGTCCGGCAGCAGAACTTCCGGCAGCGATCAAG 1140
DB	1081	AATACCATGTTGACCGCATTTACGCCCTCGTCCGGCAGCAGAACTTCCGGCAGCGATCAAG 1140
QY	1141	GCTCAACCGGGTATTGCCGATTAAGCGCGGGTAAATGGGGGAAACCTTTATCCAGTGGGTC 1200
DB	1141	GCTCAACCGGGTATTGCCGATTAAGCGCGGGTAAATGGGGGAAACCTTTATCCAGTGGGTC 1200
QY	1201	GTGGAAGATAATTTCCGTGATGTCGGTCCGGCACTCGAAGGTCGTGTCGAACCTGGTG 1260
DB	1201	GTGGAAGATAATTTCCGTGATGTCGGTCCGGCACTCGAAGGTCGTGTCGAACCTGGTG 1260
QY	1261	GGTCGGTAAATCCCTTATGAAGGCGGAAGTTTCGCAATCTTTAACTCTTACACAGTTGC 1320
DB	1261	GGTCGGTAAATCCCTTATGAAGGCGGAAGTTTCGCAATCTTTAACTCTTACACAGTTGC 1320
QY	1321	ATCGCTGGGCGAGTACGTTAATCGTCAAAATATATCCACGAAAGCAACATGACCGAT 1380
DB	1321	ATCGCTGGGCGAGTACGTTAATCGTCAAAATATATCCACGAAAGCAACATGACCGAT 1380
QY	1381	TTTATCTATCAGATTGCCGACCGCTACGTGACGGAAGATGTCTATCTTCTGCTGGCGGAT 1440
DB	1381	TTTATCTATCAGATTGCCGACCGCTACGTGACGGAAGATGTCTATCTTCTGCTGGCGGAT 1440
QY	1441	AACGGTATCGATTGTCACACCTACCGTGTATGTTGTTACTCAAGCGTTTACCACATCCACAT 1500
DB	1441	AACGGTATCGATTGTCACACCTACCGTGTATGTTGTTACTCAAGCGTTTACCACATCCACAT 1500
QY	1501	ATTCAGGACACCAACCAACCGCTCGCTGGCGATGGTTTCTCGAAATTTCCGGCGATGATT 1560
DB	1501	ATTCAGGACACCAACCAACCGCTCGCTGGCGATGGTTTCTCGAAATTTCCGGCGATGATT 1560
QY	1561	GGCCCCACACTCGGAGAGTGTACACGAGGCGGTTCGCCGCAATGCCACCGCATGTTA 1620
DB	1561	GGCCCCACACTCGGAGAGTGTACACGAGGCGGTTCGCCGCAATGCCACCGCATGTTA 1620
QY	1621	CCTGCACTCTTTTACGTTATTCATGGAGCAGTGGCATACAGGCAAACTGCCCTATGAATAT 1680
DB	1621	CCTGCACTCTTTTACGTTATTCATGGAGCAGTGGCATACAGGCAAACTGCCCTATGAATAT 1680
QY	1681	CAGGATGGCATCTTTCATGACACCGCTGTCATGCAATGTTACAGTCTGCCGATCCGCTC 1740
DB	1681	CAGGATGGCATCTTTCATGACACCGCTGTCATGCAATGTTACAGTCTGCCGATCCGCTC 1740
QY	1741	GCTGTTTATGCCAGTGAATAAGCGCTGTTTGGCGATTTTAAACGAAAGTGAAGATTTTGGC 1800
DB	1741	GCTGTTTATGCCAGTGAATAAGCGCTGTTTGGCGATTTTAAACGAAAGTGAAGATTTTGGC 1800

QY 1801 GCGTTGTTGCGGAAAAATCGCTGACGCTACCGGTTAAATTAACAA 1848
 DB 1801 GCGTTGTTGCGGAAAAATCGCTGACGCTACCGGTTAAATTAACAA 1848

RESULT 2

AAA97430
 ID AAA97430 standard; DNA; 4115 BP.

XX AAA97430;
 XX
 XX
 DT 29-JAN-2001 (first entry)

XX Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH) gene.

XX D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
 KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
 KW L-ascorbic acid biosynthesis; vitamin C; ds.

XX Gluconobacter oxydans.

XX WO200055329-A1.

XX 21-SEP-2000.

XX 16-MAR-2000; 2000WO-JP01608.

XX 17-MAR-1999; 95JP-0072810.

PR 06-AUG-1999; 95JP-0224679.

XX (FUJI) FUJISAWA PHARM CO LTD.

XX Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
 PI Yamashita M, Takata Y;

XX WPI; 2000-587530/55.

DR P-PSDB; AAB23172.

XX Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
 PT culturing its gene-transformed host cells, useful for producing
 PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 PT acid production

XX Claim 9; Page 60-63; 72pp; Japanese.

XX The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
 CC (SLDH; AAB23172) and to the gene encoding it (AAA97430). SLDH has a
 CC molecular weight of about 54 kDa and catalyses the conversion of
 CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
 CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
 CC act on xylitol, ribitol, inositol and glycerol. The invention also
 CC encompasses expression vectors and host cells comprising the
 CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
 CC The invention further relates to a method for preparing L-sorbose by
 CC contacting the recombinant SLDH with D-sorbitol; a process for producing
 CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
 CC dehydrogenase and/or sorbose dehydrogenase with L-sorbose; and a
 CC process for preparing L-ascorbic acid or its alkaline earth metals salts
 CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
 CC L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 CC acid production. The present sequence represents the Gluconobacter
 CC oxydans SLDH gene.

XX Sequence 4115 BP; 833 A; 1133 C; 1270 G; 879 T; 0 other;

Query Match 5.7%; Score 105; DB 21; Length 4115;
 Best Local Similarity 46.8%; Pred. No. 4.6e-24;
 Matches 509; Conservative 0; Mismatches 560; Indels 18; Gaps 5;

QY 503 TGCACATCGGTTAGTCTTTTCATCGCGACATCAGCGGTTATCTACACCGTTTCC 562

DB 622 TCCATTTCCGTTAGTAACTTTTTCGAGCCCATCAGCGGTTCTACGTCGACGATTC 681

QY 563 AGTGATGGCGGATAAACACGCTGGAGCAATGCTGCGGGCAATATTCGTAATGA---TGCTG 619
 DB 682 TTGAACACGCTCCGACCTGGGCGATTGTTGGTGTGGCCCTGACGGGACGTCGTTCAA 741
 QY 620 RACATGCTGCACAGCACTAGTCACAGAAAGGTCGCTATGCTGTGGAAACCGTCAGCC 679
 DB 742 AGAAAAAGCCGAGGAATTCAGGCCACGACTGCTGTATTTCCCTGACCGAGAGCGGTC 801
 QY 680 CGGAAGGGGTAAAGCAATATGAAGAGATCACCTCAATTCAGAAAGTTGATACCGTGGCAGG 739
 DB 802 CGTCCGCAAGAGACACGGTGCCTCATGGGCGCGCTGCTGCTACTATCTGCTTGGCCCGG 861
 QY 740 CAGATTTACAAACCGCTGATGCTGAAGGGGAGATCCGAAGACAAAGATGATGTTTCA 799
 DB 862 CCGATCCGGAAGCCCTGCTGAAGCATCTTGTGTGATCGGCCCATCGGTCATGTTTCCATGA 921
 QY 800 CGCTCACCAGCGGGGTACTACTACCTGA---ATACAGTACACAACTGGAAGTTAAACAATC 856
 DB 922 CGATCAGGAAGGCGGCTACAACATCAGAGACACCGGTCGCTGCTGGAATG 981
 QY 857 CTGATTTAGCGCGAGATCTTAAAGGGGATGCAAAACAAATTTACGGTGTATTATACCCGTA 916
 DB 982 CGGCAGTAAAGCCGACCTCAAGAACCCGGAAGACCGGCTACCGGTTTTCGGTTACGTGG 1041
 QY 917 TCCTCGAAGCGGTATGCAAAATAACGCCG-----ACCCTAACCCCTGCTGAATTCGG 970
 DB 1042 TCGAGGCCCTGCTGCTGCTGGGATGCGCGTGGTAAAGCATTTACGGTCATCTCCTGTG 1101
 QY 971 ATAACGTGCGCCATATGCTGAACGTTTCCATGATGGCCCTGCTGAGTTTCTCCAGCTAA 1030
 DB 1102 ATAACCTGCGTCATACCGCAATGTCGCCCGCAAGCCCTTCTCTCGGCTATGCGAAG--G 1158
 QY 1031 CTGGCAACAGGATGTCATCGACTGCTGAGTACAAATACCACTTCCCGCAATACCATGG 1090
 DB 1159 CGCGCATCCGAGTTGGCGAAGTGATGAGGAAACCGACCTTCCCGAAGCAATG 1218
 QY 1091 TTGACCGCATACGCTCTCCGCGCAGCAGAACTTCCCGCAGCGATCAAGGCTCAACCG 1150
 DB 1219 TTGATCGCATACCGCGACCGTTTCCGCGGAAATCGCAAGAGCTCAACGCGGCCAGTG 1278
 QY 1151 GTATTGCGCATAAAGCGCGGTAATGGGCAAAACCTTTATCCAGTGGGTCGTGGAAGATA 1210
 DB 1279 GCGTGATGACACCTGCGCGCTGCTGGCGAGGATTTCCATCAGTGGGTCGTGGAAGACC 1338
 QY 1211 ATTTCGCTGATGCTCGGCACTGGAGAAGTCTGGTGTGCAACTGGGTCGGGTAA 1270
 DB 1339 AGTTTCGCGATGCGCTCCGCGCTTGAAGAACCGCGCTGCAGATGCTCGGGGACGTGA 1398
 QY 1271 TCCCTATGAAGAGGGGAAGATTCGATTTCTTAACCTTTTACACACATTCGATCGCCTGG 1330
 DB 1399 CGGACTGGGATACGTTCAAGATCCGAATGCTCAATGCAGGGCATGTCATGCTCTGCTTC 1458
 QY 1331 CAGGTACGTTAATCGGTCAAAATATATCCAGAAAGCAATGACCGGATTTATCTATC 1390
 DB 1459 CAGGATTTCTGCTCGGCTATGAGAATGTGGATGACGCCATTGAAGACAGCAACTCCTTG 1518
 QY 1391 AGATTGCCACCGCTACGTGACGGAAGATGTCATTCTCTGCTT---GGCGATAACGGTA 1447
 DB 1519 GCAATCTGAAGAACTATCTCAACAGGATGTATCCCGACCTGAAGGCGCCTTCAGGCA 1578
 QY 1448 TCGATTGTCACACCTACCGTGTATGTTGTAATCAAGCGTTTTTACCAATCCACATATTTCAG 1507
 DB 1579 TGACGCTCGAAGGCTATCGGGACAGCGTCATCAGCGGTTTTCTCCAACAAGGCGATGCG 1638
 QY 1508 ACACCAACCAAGCGCTGCTGCGGATGTTTCTCGAAATATTCGGCGATGATTTGCCCCCA 1567
 DB 1639 ACCAGACGCTCGGATGTTGATGCGATGGTGTTCGAAGTTTCAGGTGTTCTGGACGAA 1698
 QY 1568 CACTGG 1574
 DB 1699 CCGTGG 1705

CC endometrial bleeding disorders, diabetic retinopathy, some forms of
 CC macula degeneration, haemangiomas, and arterial-venous malformations.
 CC They may be useful in treating deficiencies or disorders of the immune
 CC system, by activating or inhibiting the proliferation, differentiation,
 CC or mobilisation (chemotaxis) of immune cells. The etiology of these
 CC immune deficiencies or disorders may be genetic, somatic, such as
 CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
 CC toxins), or infectious. They can also be used to treat inflammatory
 CC conditions, both chronic and acute conditions. The products can also be
 CC used for detection and diagnosis. AA232002 to AA232080, and AA49503 to
 CC AA49511 represent sequences given in the exemplification of the present
 CC invention.

XX Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;

Query Match 4.7%; Score 87.6; DB 20; Length 9810;
 Best Local Similarity 48.1%; Pred. No. 6.4e-18;
 Matches 410; Conservative 0; Mismatches 424; Indels 18; Gaps 5;

Qy 503 TGCACATCGGTTAGGTTCTTTTCATCCGCGACATCAGCGGTGGTATCTACACCGTTTTCG 562
 Db 6749 TCCATATCGGCTCGGCACTTCCACCGGCGCATCAGCGGTCTATCTCGACGATCTCT 6808
 Qy 563 AGGTATGGGCGA---TAAACGCTGGACATCTGTCGGGCAATATTGTAATGATGCTG 619
 Db 6809 TCGGCTGGGCGAGGCGCACGACTGGCCATCTCTCGGCGCGGCTCTCGCCGACCGATG 6868
 Qy 620 AACATGTCGTACAGCACTAGTCACAGAAAGGTCGTATGCTGTGGAACCGTCAGCC 679
 Db 6869 CCGGATCGCGAGGCTCTGCGCGGAGGACATCTCTCGACGCTGATCGAGCTCGATC 6928
 Qy 680 CGGAAGGGGTAAAGCAATATGAAGATCACCTCAATTCAGAAAGTTGATACCGTGGCAGG 739
 Db 6929 CGGCGGCGACCGGG---CCGCGCAGGTGGGGCGATGTTGCGGCTTCTCGCGTCGAGG 6985
 Qy 740 CAGATTTACACCGCTGATGCTGAAGGGGCGAGATCCCAAGACAAAGATGATGCTTCA 799
 Db 6986 CCGAATATCGGCGCTGATGAGGCGCATGTGCGGATCCGCGCATCGCTCTCGCTGA 7045
 Qy 800 CCGTCACCAAGAGGGGTACTACCTGAATACAGTCACAAACTGGAAGTTAAACAATCTCTG 859
 Db 7046 CCGTGACGAGGGCGGCTTATGTCGATGCTCGGGCGCTTCCATCGGACGATCCCG 7105
 Qy 860 ATTTAGCGGCGAGT-----CTTAAGGGGGATGCAAAACAAATTTACGGTGTATTACCC 913
 Db 7106 ATATCGTGGCGATCGCGCCCATCTCTCGCGCGCGACGCGCTTCGGCGGATCTCTCG 7165
 Qy 914 GTATCTCGAAGCGGTATGCAATATACGCGGACCACTAACCTGCTGATTCGGATA 973
 Db 7166 CCGCCCTCCGCGCGCGCGCGCGCGGGGTTACACCTTCCACCGTATGTCCTCGGACA 7225
 Qy 974 ACGTCGCGCATATGTTCAACGTTTCCATGATGGCTGTGAGTTCTTCCAGCTAACTG 1033
 Db 7226 ACCTCCCGGCAAGCGCCATGTACCCCGCACGCGGTGGGCTGGCGAGCTCTACG 7285
 Qy 1034 GCAACAGGATGTATCGACTGGCTGAGTACAAATPACCACTTGCCCGAATACCATGGTTG 1093
 Db 7286 ACGCGGAGCTTG---CGGGCTGGGTGAAGGCGCGAGTGGCTTCCCGAAGCGCATGGTGG 7342
 Qy 1094 ACCGATTAACGCTGCTCGGCGAGCAACTTCCGGGACGATCAAGGCTCAACCGGTA 1153
 Db 7343 ACCGATCAACCCCGCCACCGCGCGCGAGCGGAACTGCGCGCAGG---GCTTCGGCGC 7399
 Qy 1154 TGGCGCATAAAGCGCGGTAAATGGCGAAACCTTTTACGAGTGGTCTGGAAGATAATT 1213
 Db 7400 TCGCCGATCCGGTGGCCCGTCACTCGAGCGCTTCCGCGAGTGGGTGATCGAGGATCAAT 7459
 Qy 1214 TCCGATGATCCGTCGCGCACTGGAGAAGGTCGGTGTGCGAACTGTGGCGTGGTAAATCC 1273
 Db 7460 TCCCGCGGAGCGCGCGCTCGAGAAGGTGGCGGTGACCTTCAACCCCGCATGTCATCG 7519
 Qy 1274 CTTATGAAGAGCGGAAGATTTCGATTTTAACTCTTTCACAGATTTGCAATCGGCTGGGAG 1333

Db 7520 CCTAGAGCGCATGAAGATCCGCATCTCTGAACGGGGGCGCATCGGTGATCGCTATCCGT 7579
 Qy 1334 GTACGTTAATCG 1345
 Db 7580 CGGCGCTCATGG 7591

RESULT 5
 AAC90082
 ID AAC90082 standard; DNA; 9810 BP.
 XX AAC90082;
 AC AAC90082;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE AF018073 cDNA clone.
 XX
 KW METH; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
 KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;
 KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;
 KW Crohn's disease; atherosclerosis; birth control; ss.
 XX Unidentified.
 OS
 XX WO200071577-A1.
 PN
 XX 30-NOV-2000.
 PD
 XX
 PF 25-MAY-2000; 2000WO-US14462.
 XX
 PR 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRETT J A.
 XX
 IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 Fornwald JA, Terrett JA;
 WPI; 2001-025136/03.
 XX
 METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 XX Claim 7; Pages 653-659; 768pp; English.
 PS
 XX The present invention relates to human METH1 and METH2, (ME for
 CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,

CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer.
XX
SQ Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;

Query Match 4.7%; Score 87.6; DB 22; Length 9810;
Best Local Similarity 48.1%; Pred. No. 6.4e-18;
Matches 410; Conservative 0; Mismatches 424; Indels 18; Gaps 5;

QY 503 TGCACATCGGGTTAGGTTCTTTTCATCGCGCACATCAGCGGTGTATCTACACCGTTTGC 562
Db 6749 TCCATATCGCGGTGCGGAACCTTCACCGCGCATCAGCGGTCTATCTCGACGATCTCT 6808

QY 563 AGTGATGGCGA--TAAAGCTGGAGCATTCGTCGGGCAATATTCGTAATGATGCTG 619
Db 6809 TCGCGCTGGCGGAGGCGCAGACTGGGCCATCTCGCGCGGGCGTCCGCCACCGATG 6868

QY 620 AACATGTCTGACAGGCACTCAGTCACAGAAAGGTCTATGTCTGGAACCGTCAGCC 679
Db 6869 CGCGGATGCGGAGGCTCTGCGCCGCGCAGCAATCTCTCGACGGTGATCGAGCTCGATC 6928

QY 680 CGGAAGGGTTAAGGAATATGAAGACATCACCTCAATTCAGAAGTTGATACCGTGGCAGG 739
Db 6929 CGCGGGCCACCGGG--CCCGCAGGTGGGGCGATGTGGGCTTCTCGCGGTGCGAGG 6985

QY 740 CAGATTACAAACCGCTGATGCTGAAGGGCGAGATCGGAAGACAAAGTATTGCTTTCA 799
Db 6986 CCGAATATCGCGCCCTGTATGAGGCCATGTCGGATCGCGCATCCGCGATCGTCTCGCTGA 7045

QY 800 CCGTCACCGAAGCGGGTACTACTGTAATACCACTGACAAAGTGAAGTTAAACAATCCTG 859
Db 7046 CCGTGACCGAGGCGGCTATTATGTCATGCTCGGGCGCTTCGATCCGACGATCCCG 7105

QY 860 ATTAGGGCGAGT-----CTTAAGGGGGATGCAAAACAATTTACGGTGTATTACCC 913
Db 7106 ATATCTGGCGGATGCGGCCCATCTCTCGCGCGCCGACCGCTTCGGCGGATCTCTCG 7165

QY 914 GTATCTCTGAAGCGGTATGCGCAATTAACCGCGGACCACTAACCTCTCAATTTGGGATA 973
Db 7166 CCGCCCTCCCGCCCGCGCGGCGGTTTACACCTTCACCGTATGTCCTGCGACA 7225

QY 974 ACGTGCGCCATAATGTGAACGTTTCCATGATGCGCTGTGTTGATTTCTCCAGCTAACTG 1033
Db 7226 ACCTCCCGGCAACGGCCATGTACCGCGAACCGCGTGGGCTTGGCGGAGCTCTACG 7285

QY 1034 GCAACAGGATGTCATCGACTGGCTGNGTACAATACCACTTCCCGCAATACCATGGTTG 1093
Db 7286 ACGCCGAGCTTG---CGGGCTGGGTGAAGCGCAGGTGGCTTCCCGAACGGGATGTCG 7342

QY 1094 ACCGATTAACGCTCTGTCGGGACGAGAACTTCCGGCAGCGATCAAGGCTCAACGGGTA 1153
Db 7343 ACCGATATCCCCCGCCGACCGCGCGGCGACGAGCGGAACCTGGGCGAG---GCTTCGGCC 7399

QY 1154 TTCCGATAAAGCGCGGTAATGGGGAAACCTTTATCCAGTGGGTGGTGGGAAGATAATT 1213
Db 7400 TCGCCGATCGGTCGCGTACCTCCGAGCGCTTCCGGCAGTGGGTGATCGAGGATCAT 7459

QY 1214 TCCGATGATTCCTCGCGGCACTCGAGAAGTTCGGTGTGCAACTGGTGGCGTGGTAAATCC 1273
Db 7460 TCCCGCGCGGACCGCGCGCTCGAGAAGGTGGGCGTGACTTCCACCCCGCATGTCCATG 7519

QY 1274 CCTATGAAGGCGAGATTCGATCTTAACCTTTCACACAGATTCGATCGCTCGGCGAG 1333
Db 7520 CCTACGAGGCGATGAAGATCCGCATCTCTGAACGGGGGCGGATCGGTTGATCGCTATCCGT 7579

QY 1334 GTACGTTAATCG 1345
Db 7580 CGCGGCTCATGG 7591

RESULT 6

AAS92759/c

ID AAS92759 standard; cDNA; 921 BP.

XX AAS92759;

AC AAS92759;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #28563.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR P-PSDB; ABG28572.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 1; SEQ ID No 28563; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 921 BP; 230 A; 235 C; 230 G; 226 T; 0 other;

Query Match 4.4%; Score 80.6; DB 23; Length 921;

Best Local Similarity 50.7%; Pred. No. 3.5e-16;

Matches 194; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

QY 1169 CGGTAATGGCGCAACCTTTATCCAGTGGTCTGGGAAGATAATTTCCGATGTCGCTC 1288

Db 917 CGATTAGTCGGAACCGTTTATCCAGTGGTCTGGGAAGATAACTTCTCCGTCGGCGTC 858

QY 1229 CGGCACTGGGAAGGTGGTGTGCAACTGGTGGCGTAAATCCCTATGACAGAGCGGA 1288

Db	439	GTGGGTGGTGGAAAGATAA	CTTCGCTGGCGGTCTCGCTGGGAAGTCGCAGGTGTACA	380
Qy	1254	ACTGGTGGCGTGGTAATCC	CTATGAAGAGCGGAGATTCCGATTTCTTA	1313
Db	379	AATGGTGAATGATGTCTG	CGCATGGGAAGAGATGA	320
Qy	1314	CAGTTGCATCCGCTGGG	CAGGTACGTTAATCGTCAAAAATATATCCACGAAAGCACAA	1373
Db	319	CTCTTTCTCGCTATATCT	GGGTACCTTCAGGATTCGCCATATCACTGATTCGATGCA	260
Qy	1374	GACGATTTATCTATCAG	ATTCGACGCGCTACGTACGGAAGATGTCATTCCTGCTT	1433
Db	259	GGATCGCGCAATTTGCG	CCATCGCCAGAACATTAATGCTGGATGACGAAGCGCGACACT	200
Qy	1434	GGCGGATAACGGTATCG	ANTTCCCAACCTTACCGTGATGTTGCTACTCAAGCGTTTTACCA	1493
Db	199	GCAAAATTAAGATGTG	ATTAACAAATATCGGATTAAGTTAATTCACCGTTTTGCTAA	140
Qy	1494	TCCACATATTTCAGGAC	CAACCAACGCGCTCGCTGCGGATGTTTCTCGAAAATTCGG	1551
Db	139	TCCGGCGCTGAACATA	GAACCTGGCAATCCGATGATGCGACCGACAATTAACG	82
RESULT 12				
AAFL12753				
ID	AAF12753 standard; cdNA; 684 BP.			
AC	AAF12753;			
XX	13-MAR-2001 (first entry)			
DE	Aspergillus oryzae EST SEQ ID NO:5276.			
XX	Multiple gene expression; filamentous fungal cell; EST;			
KW	expressed sequence tag; Fusarium venenatum; Aspergillus niger;			
KW	Aspergillus oryzae; Trichoderma reesei; identification; recombination;			
KW	culture condition; environmental stress; spore morphogenesis;			
KW	metabolic pathway engineering; catabolic pathway engineering; ss.			
XX	Aspergillus oryzae.			
OS	WO200056762-A2.			
PN	28-SEP-2000.			
PD	22-MAR-2000; 2000WO-US07781.			
XX	22-MAR-1999; 99US-0273623.			
PF	(NOVO) NOVO NORDISK BIOTECH INC.			
PR	(NOVO) NOVO NORDISK AS.			
PA	Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;			
PI	WPI; 2000-594572/56.			
DR	Monitoring differential expression of genes in filamentous fungal cells			
XX	uses fluorescence-labeled nucleic acids isolated from the cells and a			
PT	substrate of expressed sequence tags.			
PT	Claim 88; Page 2202-2203; 3161pp; English.			
PS	The present invention describes a method for monitoring differential			
CC	expression of genes in a first filamentous fungal (FF) cell relative to			
CC	expression of the same genes in one or more second filamentous fungal			
CC	cells. The method uses fluorescence-labeled nucleic acids isolated from			
CC	the FF cells and a substrate of expressed sequence tags (EST). The ESTs			
CC	are used in the methods for monitoring differential expression of genes			
CC	in a first filamentous fungal (FF) cell relative to expression of the			
CC	same genes in one or more second filamentous fungal cells. Monitoring			
CC	the global expression of genes from FF cells allows the production			
CC	potential of the microorganisms to be improved. New genes may be			

CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

XX Sequence 684 BP; 152 A; 212 C; 185 G; 135 T; 0 other;

Query Match 3.3%; Score 61.2; DB 21; Length 684;
Best Local Similarity 54.4%; Pred. No. 1.1e-09;
Matches 123; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 1034 GCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTTCCCGAATACCATGTTG 1093

Db 8 GCACCCCGAGATCGGAATGGATTCGCCGACAGGGCGCTTCCCAACCCCATGGTCG 67

Qy 1094 ACCGCATTACGCCCTCGTCGCGCAGCAGACAACTTCGCGCAGCGATCAAGGCTCAACGGGTA 1153

Db 68 ACCGTATCACCCCTCAGACATCCGCCACCGGATAAACGGCACTCGCAGACAACTTGCCA 127

Qy 1154 TTCCCGATAAGCCCGGTAAATGGCGGAACCTTTATCCAGTGGTGTGGAAGATAATT 1213

Db 128 TCGAGGACTCGTGGCCGCTGTCACAGAGCCCTTTATGCAGTGGGTAATTGAGGATCAGT 187

Qy 1214 TCCGTGATGTCGTCGCGCACTGAGAGGTCGCTGTCGAACCTGGT 1259

Db 188 TCTCCGATGGCCGCCACCATTCAGAGAGGTGTGTCGCCAGTGGT 233

RESULT 13

AA82343
ID AAS82343 standard; cDNA; 2193 BP.

XX AC AAS82343;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #18147.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG18156.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

XX PS

Claim 1; SEQ ID NO 18147; 103pp; English.

XX CC

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.

XX CC

Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pcc_sequences.

XX SQ

Sequence 2193 BP; 494 A; 592 C; 603 G; 504 T; 0 other;

Query Match 2.3%; Score 43.4; DB 23; Length 2193;
Best Local Similarity 66.7%; Pred. No. 0.0025;

Matches 62; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1168 CCGTAATGGCGGAACCTTTATCCAGTGGTGTGGAAGATAATTCGTCGATCCGT 1227

Db 799 CAGATGACTGTGCAACCGTTATCCAGTGGTGTGGAAGATAACTTCGTCGTCGGCGT 858

Qy 1228 CCGCACTGGAGAGTGGTGTGCAACTGGTG 1260

Db 859 CTGCTGGGAAGTCGCAGGTGTACAAATGGTG 891

RESULT 14

AAH48024

ID AAH48024 standard; DNA; 14041 BP.

XX AC AAH48024;

XX DT 18-SEP-2001 (first entry)

XX DE Internal control B19c #1.

XX KW Internal control; ss.

XX OS Parvovirus.

XX PN WO200146463-A2.

XX PD 28-JUN-2001.

XX PF 20-DEC-2000; 2000WO-EPI2996.

XX PR 22-DEC-1999; 99AT-0002170.

XX PA (BAXT) BAXTER AG.

XX PI Zimmermann-K, Turecek P, Schwarz H, Rieger M;

XX DR WPI; 2001-408658/43.

XX PT Internal standards useful for nucleic acid amplification assays,
XX PT comprises a synthetic nucleic acid made by non-recombinant techniques

XX Example 1; Page 22-27; 30pp; English.

XX The present invention relates to methods for the preparation and use of
CC internal controls for nucleic acid amplification assays. The internal
CC controls comprise a synthetic nucleic acid made by non-recombinant
CC techniques. The internal controls are useful for detecting nucleic acids
CC in a sample such as blood, spinal fluid, semen, saliva, tears, cell
CC culture fluid, recombinant cells, animal tissue or plant tissue, by a
CC quantitative PCR assay, by adding the internal control to the sample,
CC amplifying the nucleic acids in the sample and detecting the amplified
CC products. The internal controls help in performing the nucleic acid
CC amplification assay quickly and inexpensively without sacrificing assay
CC specificity or sensitivity. The present sequence is one such internal
CC control, which was used in an example to illustrate the present
CC invention.
CC Note: the present sequence is the SEQ ID 8 shown in the sequence listing.
CC This sequence differs from the SEQ ID 8 shown on page 13 of the
CC disclosure (see AAH48038).
XX
SQ Sequence 14041 BP; 2151 A; 1354 C; 466 G; 2218 T; 7852 other;
Query Match 2.3%; Score 42.4; DB 22; Length 14041;
Best Local Similarity 18.8%; Pred. No. 0.017;
Matches 72; Conservative 132; Mismatches 177; Indels 3; Gaps 1;
QY 1285 GCGAAGATTCGCAATCTTAACCTTCCACACAGTTGCATCCCTGGCAGGTACGTTAATC 1344
DB 2734 SSSCHGNCDSWTHNANRMYSHRTTMANDATVRYCNCSTASARSCADGNCDSYNTH 2793
QY 1345 GGTCAAAATATATCCAGCAAGACCAATGACCGATTTCATCTATCAGATTGCGGACCGC 1404
DB 2794 SRSSCHASRAMGSYNTHSSRWGBCUAGBNBGRGMANYTHANTCATNTHGNCDSYN 2853
QY 1405 TAGTCGACGGAAGATGTCATCTCTCTGGCGGATAACGGTATCGATTTCGCAACCTAC 1464
DB 2854 THSDSAYTAKSACATMATCAYNTHSYNTHSRRTSRADYRVDBYTHSYNTHSRYSYTMN 2913
QY 1465 CGTGATGTTGCTACTCAAGCGTTTACCAATCCACATATTCAGCACACCAACCGGTC 1524
DB 2914 ADD---TNTHSANTCATNSNTNRRDBYTHRSNCRNTHWCHCANBWNATDNYASARS 2970
QY 1525 GTCGCGGATGTTCTCGAAATTCGCGGATGATTGCCGCCACACTGGGAGAGTGCTAC 1584
DB 2971 MCNSMNGNSVRCATNANDTRACTNRCSSNTHCAHBGNCRARATNSTANDARDSNTHR 3030
QY 1585 CACGAGGCGTTCGCCGGAATGCCAGCGCATGTTACCTGCACTGTTTACGTTATTCATG 1644
DB 3031 RARTSRNSGYTHRCSSNACCRDANCWTHNVTNNSRBYWSTDTARGSCANDSTRASA 3090
QY 1645 GAGCAGTGGCATCACGGCAAACTG 1668
DB 3091 NDCMRHNSVMDCASCRNNGTSTS 3114
RESULT 15
AAS92760
ID AAS92760 standard; cDNA; 2238 BP.
XX
AC AAS92760;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28564.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Llu C, Tang YT;
PI
PI WPI: 2001-639362/73.
DR P-PSDB; ABC28573.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 28564; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2238 BP; 600 A; 491 C; 577 G; 570 T; 0 other;
Query Match 2.1%; Score 38.4; DB 23; Length 2238;
Best Local Similarity 45.6%; Pred. No. 0.13;
Matches 135; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
QY 1256 TGTGGCGTCGGTAATCCCTATGAAGAGCGGAGATTCGCATTCTTAACCTTTCACACA 1315
DB 2 TGTGTAATGATGTCCTGCCATGGGAAGAGATGAACCTGGGATGCTTATGGCAGCCACT 61
QY 1316 GTTGCAATCGCTGGCAGGTACGTTAATCGGTCAAAATATATATCCACGAAGCAATGA 1375
DB 62 CTTTCTCGCTTATCTGGGTTACCTCTCAGGATTCGCCCATATCATGATTCATGCAGG 121
QY 1376 CCGATTTTATCTATCAGATTGGGACCGCTACGTGACGGAAGATGTCATCTCTGTTGG 1435
DB 122 ATCGCGCATTTTCGCCATCGCCGCAACATTAATGCTGGATGAGCAAGCGCCGACACTGC 181
QY 1436 GCGATAACGGTATCGATTTGCCAACCTACCGTGTCTTACTCAAGCGTTTTTACCAATC 1495
DB 182 AAATTAAGATGTCGATTTAACAACATATATGCGGATGAAGTTAATGACGCTTTTGCCTAATC 241
QY 1496 CACATATTCCAGGACACCAACCGCTCGCTGCGGATGTTTCTCGAAATTCG 1551
DB 242 CGGCGCTGAACATAAGACCTTGGCAATCCGATGATGCGCAGCCAGCAAAATACCG 297

Search completed: March 21, 2003, 20:48:30
Job time : 805.996 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 09:44:37 ; Search time 67.0095 Seconds
(without alignments)
8457.593 Million cell updates/sec

Title: US-09-802-208B-1
Perfect score: 1848
Sequence: 1 atgaacgaacaattacatg.....tctacgcgtaataactaa 1848

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44.4	2.4	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	33.2	1.8	4797	4 US-09-643-597-134	Sequence 134, App
C 3	32.4	1.8	1560	4 US-09-453-702B-264	Sequence 264, App
C 4	32.4	1.8	2230	4 US-09-342-647-1	Sequence 1, Appl
C 5	32.4	1.8	3271	4 US-09-770-170-3	Sequence 3, Appl
C 6	32.4	1.8	61663	4 US-09-453-702B-62	Sequence 62, Appl
C 7	32	1.7	1407	1 US-08-459-287-1	Sequence 1, Appl
C 8	31	1.7	2796	2 US-08-937-931-3	Sequence 3, Appl
C 9	31	1.7	2796	4 US-09-285-502-3	Sequence 3, Appl
C 10	31	1.7	2796	4 US-09-709-126-3	Sequence 3, Appl
C 11	31	1.7	2796	4 US-09-871-385A-3	Sequence 3, Appl
C 12	31	1.7	3349	2 US-08-920-234-1	Sequence 1, Appl
C 13	31	1.7	3410	4 US-09-527-154-3	Sequence 3, Appl
C 14	30.8	1.7	1339	1 US-07-936-163-2	Sequence 2, Appl
C 15	30.6	1.7	699	4 US-08-998-416-591	Sequence 591, App
C 16	30.6	1.7	3200	1 US-08-444-405-1	Sequence 1, Appl
C 17	30.6	1.7	3200	1 US-08-384-850-1	Sequence 1, Appl
C 18	30.4	1.6	541	4 US-09-221-017B-68	Sequence 68, Appl
C 19	30.2	1.6	4411529	4 US-09-103-840A-1	Sequence 1, Appl
C 20	30	1.6	1240	1 US-08-103-998-1	Sequence 1, Appl
C 21	30	1.6	1661	1 US-08-518-474-1	Sequence 1, Appl
C 22	29.8	1.6	216	1 US-07-663-413-28	Sequence 28, Appl
C 23	29.8	1.6	216	1 US-08-055-530-28	Sequence 28, Appl
C 24	29.8	1.6	216	1 US-08-247-475-47	Sequence 47, Appl
C 25	29.8	1.6	216	1 US-08-479-650-47	Sequence 47, Appl
C 26	29.8	1.6	216	1 US-08-191-866D-75	Sequence 75, Appl
C 27	29.8	1.6	216	1 US-08-674-169-47	Sequence 47, Appl

C 28	29.8	1.6	216	2 US-08-185-949B-75	Sequence 75, Appl
C 29	29.8	1.6	1172	2 US-08-070-301-21	Sequence 21, Appl
C 30	29.8	1.6	1400	2 US-08-305-764C-57	Sequence 57, Appl
C 31	29.8	1.6	1400	2 US-08-305-764C-59	Sequence 59, Appl
C 32	29.8	1.6	2095	2 US-08-305-764C-55	Sequence 55, Appl
C 33	29.8	1.6	3226	2 US-08-070-301-10	Sequence 10, Appl
C 34	29.8	1.6	4403765	4 US-09-103-840A-2	Sequence 2, Appl
C 35	29.8	1.6	4411529	4 US-09-103-840A-1	Sequence 1, Appl
C 36	29.6	1.6	1370	4 US-09-221-017B-716	Sequence 716, App
C 37	29.6	1.6	1436	4 US-08-961-527-365	Sequence 365, App
C 38	29.6	1.6	4483	4 US-08-961-527-363	Sequence 363, App
C 39	29.6	1.6	11717	1 US-08-801-263A-4	Sequence 4, Appl
C 40	29.6	1.6	11717	3 US-09-102-248-4	Sequence 4, Appl
C 41	29.6	1.6	32768	4 US-08-961-527-71	Sequence 71, Appl
C 42	29.4	1.6	1550	4 US-08-858-207A-74	Sequence 74, Appl
C 43	29.4	1.6	2625	4 US-09-030-335-8	Sequence 8, Appl
C 44	29.4	1.6	8937	2 US-08-449-933-1	Sequence 1, Appl
C 45	29.4	1.6	8937	4 US-07-966-049A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F1s
US-08-232-463-14

Query Match 2.4% Score 44.4; DB 1; Length 7218;

Qy	1074	TTGCCCGCAATACCATGGTTAGCCGCAATTAAGCCTCGTCCGGCGAGCAGAACTTCCGGCAGC	1133
Db	3160	TGGCTGGAAGATCCTCGGGGTGGGGCTTGGGGCTCACACACCTGTAGCACTTACTGGTAG	3219
Qy	1134	GATCAAGCGTCAAAACGGATTTCGCCGATAAAGCGCGGTAATGGGCGAAACCTTTATCCA	1199
Db	3220	GACCAAGCATCTTTGGGGGGTGCCCGCTGAGTGGCAGGGGACAGGAGTCACTTTCTTTTCG	3279
Qy	1194	GTGGCGTGTGGAAGATAATTTCCGTGATGT 1223	
Db	3280	TGGGAGGTCTAATCTAGATATCGACTTCT 3309	
RESULT 3			
US-09-453-702B-264/c			
; Sequence 264, Application US/09453702B			
; Patent No. 6365723			
; GENERAL INFORMATION:			
; APPLICANT: Blattner, Frederick R.			
; Burland, Valerie			
; Perna, Nicole T.			
; Plunkett, Guy			
; Welch, Rod			
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157			
; NUMBER OF SEQUENCES: 265			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Quarles & Brady			
; STREET: 1 South Pinckney Street			
; CITY: Madison			
; STATE: WI			
; COUNTRY: US			
; ZIP: 53701-2113			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44mb storage			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Word Perfect 8.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/453,702B			
; FILING DATE: 03-Dec-1999			
; CLASSIFICATION: <Unknown>			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 60/110,955			
; FILING DATE: 04-DEC-1998			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Seay, Nicholas J.			
; REGISTRATION NUMBER: 27386			
; REFERENCE/DOCKET NUMBER: 960296.95017			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (608) 251-5000			
; TELEFAX: (608) 251-9166			
; INFORMATION FOR SEQ ID NO: 264:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 1560 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; SEQUENCE DESCRIPTION: SEQ ID NO: 264:			
US-09-453-702B-264			
Query Match 1.8%; Score 32.4; DB 4; Length 1560;			
Best Local Similarity 62.2%; Pred. No. 1.4;			
Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps			
Qy	765	AGGGCGAGATCCGAACACAAAAGTGATTCCTTACCGCTCACCGAAGCGGGTACTACCT 824	
Db	225	ACGCACAGTCCGTAGACAGCAGCTCTTCCTTACCGTACCAGAACCGCATACTCCCC 165	
Qy	825	GAATACAGTCAAAACTGGAA 846	
Db	165	CGGATGCATCGTCATTCGCGTAA 144	

```

Db 2230 GATCCAGTATCTTCTGTTCTGCAGATAGTTGACGGCATCGGCCACCTGACGAAAAATGT 2171
QY 1071 CACTTGCCCG 1080
    ||| ||
Db 2170 AGCTTCCAG 2161

RESULT 6
US-09-453-702B-62/c
; Sequence 62, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; plunkett, Guy
; Welch, Rod
;
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESS: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61663
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
;
US-09-453-702B-62

Query Match 1.8%; Score 32.4; DB 4; Length 61663;
Best Local Similarity 62.2%; Pred. No.15;
Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps

QY 285 AGGGGCGATCCGAGACAAAAGTGATTGCTTTTCCACCGTCCACCGAAGCGGGTACTACTCT 344
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38189 ACGCACAGCTCCGTAGACAGCAGTCTTCCCTTTACCGTCACCGAAACCGCATACTCCCC 38130

QY 345 GAATACCAAGTCACAACTGGAA 366
    ||| || || ||
Db 38129 CGGATCCATCGTCATTCGGTAA 38108

RESULT 7
US-08-459-287-1
; Sequence 1, Application US/08459287
; Patent No. 5686596
; GENERAL INFORMATION:

```



```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,931
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2796 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-285-502-3

Query Match      1.7%; Score 31; DB 4; Length 2796;
Best Local Similarity 45.9%; Pred. No. 6.1;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1264 TCGGTAATCCCTATGAAGAGCGGAGATTCCGATCTTAACTCTTCACACAGTTGCGATC 1323
DB 913 TCCGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAAGTATTTTTC 854
QY 1324 GCCTGGGAGGAGTAAATCGGTCAAAAATATATCCACGAAAGCAATGACCGATTTT 1383
DB 853 AGCTAAATTTGACGTTTTTCTCAGAAAGTTCTGGACCATTTAGCAGCATGTTCTCTTG 794
QY 1384 ATCTATCAGATTCGCGACCGCTACGTGACGGAAGATGATCTCTTGGTGGCGGATAAC 1443
DB 793 AGGTATCTGTGTTACTCTCTACACCATCTGATGTTTCTTCTTCTTCTTCTTCTTCTT 734
QY 1444 GGTATCGATTTGCCAACCTACCGTGATGTTGTTACTCAAGCGTTTACCAT 1494
DB 733 TGAATGATCTGCACAGCCCCCTGAGGACCGTATTTATGGGATAGTTAAT 683

RESULT 10
US-09-709-126-3/c
; Sequence 3, Application US/09709126
; Patent No. 6319704
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; Pan, Duojia
; Rooke, Jenny
; Yavari, Reza
; Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 6319704el Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/709,126
; FILING DATE: 08-NO. 6319704-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/285,502
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627

```

```

; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2796 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-709-126-3

Query Match      1.7%; Score 31; DB 4; Length 2796;
Best Local Similarity 45.9%; Pred. No. 6.1;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1264 TCGGTAATCCCTATGAAGAGCGGAGATTCCGATCTTAACTCTTCACACAGTTGCGATC 1323
DB 913 TCCGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAAGTATTTTTC 854
QY 1324 GCCTGGGAGGAGTAAATCGGTCAAAAATATATCCACGAAAGCAATGACCGATTTT 1383
DB 853 AGCTAAATTTGACGTTTTTCTCAGAAAGTTCTGGACCATTTAGCAGCATGTTCTCTTG 794
QY 1384 ATCTATCAGATTCGCGACCGCTACGTGACGGAAGATGATCTCTTGGTGGCGGATAAC 1443
DB 793 AGGTATCTGTGTTACTCTCTACACCATCTGATGTTTCTTCTTCTTCTTCTTCTTCTT 734
QY 1444 GGTATCGATTTGCCAACCTACCGTGATGTTGTTACTCAAGCGTTTACCAT 1494
DB 733 TGAATGATCTGCACAGCCCCCTGAGGACCGTATTTATGGGATAGTTAAT 683

RESULT 11
US-09-871-385A-3/c
; Sequence 3, Application US/09871385A
; Patent No. 6399350
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; Pan, Duojia
; Rooke, Jenny
; Yavari, Reza
; Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 6399350el Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/871,385A
; FILING DATE: 31-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/709,126
; FILING DATE: 08-NO. 6399350-2000
; CLASSIFICATION: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627

```

; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2796 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-871-385A-3

Query Match 1.7%; Score 31; DB 4; Length 2796;
Best Local Similarity 45.9%; Pred. No. 6.1;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 1264 TCGGTAATCCCTATGAAGAGCGGAGATTTCGCATCTTTAACTCTTCACACAGTTGCATC 1323
Db 913 TCGGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAGTATTTTTTC 854
QY 1324 GCCTGGCAGGTACGTTAAATCGGTCAAAAATATATCCAGAAAGCACAAATGACCGATTTT 1383
Db 853 AGCTGAATTTGTAGTTTTTCTCAGAAGTTCTGGACCATTAGCAGCATGTTCTTCTTG 794
QY 1384 ATCTATCAGATTGCCAGCCGCTACGTGACGGAAGATGTCATTCCTTGGTGGCGGATAAC 1443
Db 793 AGGTATCTGTGTTACTTCCCTACACAGTCATCTGGTATTTTCCCTCATCTTTCAATAC 734
QY 1444 GGTATCGATTGGCAACCTACCGTGTGTGTACTCAAGCGTTTACCAT 1494
Db 733 TGAATGATCTGCACAGCCCTCGAGACCGTATTTATGGGGATAGTTAAT 683

RESULT 12
US-08-920-234-1/c
; Sequence 1, Application US/08920234
; Patent No. 5922546
; GENERAL INFORMATION:
; APPLICANT: ELLIS, CATHERINE
; APPLICANT: JACKSON, JEFFREY
; APPLICANT: MAYER, RUTH
; TITLE OF INVENTION: HUMAN DISINTEGRIN METALLOPROTEASE
; TITLE OF INVENTION: RELATED TO DROSOPHILA KUZ GENE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,234
; FILING DATE: 25-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET INFORMATION: GH-70237
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-920-234-1

Query Match 1.7%; Score 31; DB 2; Length 3349;
Best Local Similarity 45.9%; Pred. No. 6.8;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 1264 TCGGTAATCCCTATGAAGAGCGGAGATTTCGCATCTTTAACTCTTCACACAGTTGCATC 1323
Db 823 TCGGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAGTATTTTTTC 764
QY 1324 GCCTGGCAGGTACGTTAAATCGGTCAAAAATATATCCAGAAAGCACAAATGACCGATTTT 1383
Db 763 AGCTGAAGTTGTAGTTTTTCTCAGAAGTTCTGGACCATTAGCAGCATGTTCTTCTTG 704
QY 1384 ATCTATCAGATTGCCAGCCGCTACGTGACGGAAGATGTCATTCCTTGGCGGATAAC 1443
Db 703 AGGTATCTGTGTTACTTCCCTACACAGTCATCTGGTATTTTCTCATCTTTCAATAC 644
QY 1444 GGTATCGATTGGCAACCTACCGTGTGTGTACTCAAGCGTTTACCAT 1494
Db 643 TGAATGATCTGCACAGCCCTCGAGACCGTATTTATGGGGATAGTTAAT 593

RESULT 13
US-09-527-154-3/c
; Sequence 3, Application US/09527154
; Patent No. 6228648
; GENERAL INFORMATION:
; APPLICANT: Thomas P. Condon
; APPLICANT: Shin Cheng Fluoroy
; TITLE OF INVENTION: ANTISENSE MODULATION OF ADAM10 EXPRESSION
; FILE REFERENCE: ISPH-0446
; CURRENT APPLICATION NUMBER: US/09/527,154
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 3
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (470)...(2716)
US-09-527-154-3

Query Match 1.7%; Score 31; DB 4; Length 3410;
Best Local Similarity 45.9%; Pred. No. 6.9;
Matches 106; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 1264 TCGGTAATCCCTATGAAGAGCGGAGATTTCGCATCTTTAACTCTTCACACAGTTGCATC 1323
Db 1180 TCGGTAATATTTAAAGAACAAATGATCAGTCTGAATATAAAGCTGACAAGTATTTTTTC 1121
QY 1324 GCCTGGCAGGTACGTTAAATCGGTCAAAAATATATCCAGAAAGCACAAATGACCGATTTT 1383
Db 1120 AGCTGAAGTTGTAGTTTTTCTCAGAAGTTCTGGACCATTAGCAGCATGTTCTTCTTG 1061
QY 1384 ATCTATCAGATTGCCAGCCGCTACGTGACGGAAGATGTCATTCCTTGGCGGATAAC 1443
Db 1060 AGGTATCTGTGTTACTTCTCTACACAGTCATCTGGTATTTCTCATCTTTCAATAC 1001
QY 1444 GGTATCGATTGGCAACCTACCGTGTGTGTACTCAAGCGTTTACCAT 1494
Db 1000 TGAATGATCTGCACAGCCCTCGAGACCGTATTTATGGGGATAGTTAAT 950

RESULT 14

US-07-936-163-2/c
; Sequence 2, Application US/07936163
; Patent No. 5743477

GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: MERLO, DONALD J
; APPLICANT: HOUTCHENS, ROBERT A
; APPLICANT: STRICKLAND, JAMES A
; APPLICANT: ORR, GREGORY L
; TITLE OF INVENTION: INSECTICIDAL PROTEINS AND METHOD FOR
; TITLE OF INVENTION: PLANT PROTECTION
; NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:
; ADDRESS: THOMAS D. ZINDRICK
; STREET: 9002 PURDUE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US

ZIP: 46268-1189
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/936,163
; FILING DATE: 27-AUG-1992
; CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
; NAME: ZINDRICK, THOMAS D
; REGISTRATION NUMBER: 32,185
; REFERENCE/DOCKET NUMBER: C-38,424A

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 636-1869

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
; LENGTH: 1339 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-07-936-163-2

Query Match 1.7%; Score 30.8; DB 1; Length 1339;
Best Local Similarity 57.1%; Pred. No. 4.4;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 429 TACCCTGTCCTGCGGCTATGCGAATAACCGCGGACCACTAACCTGATGAACA 488

Db 701 TGCAGCTCTGCTGCGGCTAACCGGATTAACCGCGGATCAGCAACAGTAGCAACAGC 642

QY 489 ACAATTTACATGGCTGCACATCGGCTAGGTTCTTTTC 526

Db 641 ACCATCAACAGATTGAATCATATTATCTCCATTAC 604

RESULT 15

US-08-998-416-591/c
; Sequence 591, Application US/08998416
; Patent No. 6239264

GENERAL INFORMATION:
; APPLICANT: Philippssen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6239264artis Corporation

STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 591:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: PAG1406RP

US-08-998-416-591

Query Match 1.7%; Score 30.6; DB 4; Length 699;
Best Local Similarity 51.1%; Pred. No. 3.4;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 945 CGGACCACTAACCCCTGCTGAATTGCGATAAATGCGCCATAATGTTGAACGTTTCCATGA 1004

Db 542 CGGTACTCTATCTCTCTGTAATGATATTAACGTCTGCCACACGAGTATATGATGG 483

QY 1005 TGGCCTGGTTGAGTTTCTCCAGCTAACTGGCAACAGGATGTATCGACTGGCTGAGTAC 1064

Db 482 TGCCACTATTACTATGACAGCTGCTGCTGCCACGACACGCGCATGAAGTGGTGCAGCAC 423

QY 1065 AAATACCACTTGCCTGCAATAC 1085

Db 422 CAATCCACGTACGCCACCAC 402

Search completed: March 21, 2003, 11:12:01
Job time : 166.009 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:05:17 ; Search time 138.147 Seconds
(without alignments)

10373.549 Million cell updates/sec

Title: US-09-802-208B-1

Perfect score: 1848

Sequence: 1 atgaacgaacaattacatg.....tctacgcgttaataactaa 1848

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1848	100.0	1848	9	US-09-802-208B-1
2	68.8	3.7	1509	9	Sequence 1, Appli
3	36	1.9	653	9	Sequence 126, App
4	33.2	1.8	671	9	Sequence 402, App
5	33.2	1.8	4797	10	Sequence 346, App
6	33.2	1.8	4797	10	Sequence 134, App
7	33.2	1.8	4797	10	Sequence 134, App
8	33.2	1.8	4797	10	Sequence 134, App
9	32.8	1.8	596	9	Sequence 310, App
10	32.8	1.8	615	10	Sequence 4050, Ap
11	32.8	1.8	919	9	Sequence 258, App
12	32.4	1.8	1560	9	Sequence 264, App
13	32.4	1.8	2436	10	Sequence 7778, Ap
14	32.4	1.8	61663	9	Sequence 62, Appl
15	32	1.7	2139	10	Sequence 9961, Ap
16	31.6	1.7	1200	10	Sequence 3, Appli
17	31.6	1.7	1833	10	Sequence 18, Appl
18	31.6	1.7	1833	10	Sequence 21, Appl
19	31.4	1.7	2176	9	Sequence 8, Appli

20	31.4	1.7	2265	9	US-09-843-250-1	Sequence 1, Appli
21	31.4	1.7	2265	9	US-09-843-250-25	Sequence 25, Appl
22	31.4	1.7	2265	9	US-09-843-250-27	Sequence 27, Appl
23	31.4	1.7	2265	9	US-09-843-250-28	Sequence 28, Appl
24	31.4	1.7	2265	9	US-09-843-250-29	Sequence 29, Appl
25	31.4	1.7	2265	9	US-09-843-250-30	Sequence 30, Appl
26	31.4	1.7	2265	9	US-09-843-250-31	Sequence 31, Appl
27	31.4	1.7	2265	9	US-09-843-250-56	Sequence 56, Appl
28	31.4	1.7	2265	9	US-09-843-250-57	Sequence 57, Appl
29	31.4	1.7	2294	9	US-09-843-250-6	Sequence 6, Appli
30	31.4	1.7	2515	9	US-09-843-250-4	Sequence 4, Appli
31	31.4	1.7	4355	9	US-09-843-250-7	Sequence 7, Appli
32	31.4	1.7	4744	10	US-09-775-938A-32	Sequence 32, Appl
33	31.4	1.7	9706	9	US-09-843-250-5	Sequence 5, Appli
34	31.4	1.7	9841	9	US-09-843-250-3	Sequence 9, Appli
35	31.4	1.7	14462	9	US-09-843-250-9	Sequence 2430, Ap
36	31.2	1.7	3262	10	US-09-764-869-2430	Sequence 2433, Ap
37	31.2	1.7	3262	10	US-09-764-869-2433	Sequence 841, App
38	31	1.7	341	10	US-09-770-791-841	Sequence 866, App
39	31	1.7	341	10	US-09-924-035A-866	Sequence 1409, App
40	31	1.7	783	9	US-09-938-842A-1409	Sequence 849, App
41	31	1.7	1800	9	US-10-101-464A-849	Sequence 187, App
42	31	1.7	2594	10	US-09-070-927A-187	Sequence 3, Appli
43	31	1.7	2796	10	US-09-871-388-3	Sequence 2460, Ap
44	30.8	1.7	521	10	US-09-815-242-2460	Sequence 4674, Ap
45	30.8	1.7	741	10	US-09-974-300-4674	

ALIGNMENTS

RESULT 1

US-09-802-208B-1
; Sequence 1, Application US/09802208B
; Publication No. US20030041352A1
; GENERAL INFORMATION:
; APPLICANT: Parrott, Wayne
; APPLICANT: LaFayette, Peter
; APPLICANT: Kane, Patrick
; TITLE OF INVENTION: Arabitol or Ribitol As Positive Selectable Markers
; FILE REFERENCE: UGA-855R
; CURRENT APPLICATION NUMBER: US/09/802.208B
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-802-208B-1

Query Match	100.0%	Score	1848	DB	9	Length	1848	
Best Local Similarity	100.0%	Pred. No.	0	Mismatches	0	Indels	0	Gaps
Matches	1848	Conservative	0					
Qy	1	ATGAACGAACAATTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGGCACATCAG	60					
Db	1	ATGAACGAACAATTACATGGCTGCACATCGGGTTAGGTTCTTTTCATCGGCACATCAG	60					
Qy	61	CGGTGGTATCTACACCGTTTCAGGTGATGGCGGATAAACGCTGGAGCATTCGTCGGGC	120					
Db	61	CGGTGGTATCTACACCGTTTCAGGTGATGGCGGATAAACGCTGGAGCATTCGTCGGGC	120					
Qy	121	AATATTCGTAATGATGCTGAACATGCTGACAGGCACCTCAGTGCACAGAAAGTGCCTAT	180					
Db	121	AATATTCGTAATGATGCTGAACATGCTGACAGGCACCTCAGTGCACAGAAAGTGCCTAT	180					
Qy	181	GTGCTGGAACCGTCAGCCCGAAGGGTAAAGCAATATGAAGAGATCACCTCAATTTCAG	240					
Db	181	GTGCTGGAACCGTCAGCCCGAAGGGTAAAGCAATATGAAGAGATCACCTCAATTTCAG	240					
Qy	241	AAGTTGATACCGTGGCAGGCAGATTTACACCGCTGATTGCTGAAGGGGCAGATCCGAAG	300					
Db	241	AAGTTGATACCGTGGCAGGCAGATTTACACCGCTGATTGCTGAAGGGGCAGATCCGAAG	300					

Qy	1381	TTTATCTATCAGATTCGCCGACCGCTACGTGACGGAGATGCTATTCCTTGCCTTGGCGGAT	1440
Db	1381	TTTATCTATCAGATTCGCCACCGCTACGTGACGGAGATGCTATTCCTTGCCTTGGCGGAT	1440
Qy	1441	AACGGTATCGATTTCGCAACCTACCGTGATGTTGTACTCAAGCGTTTTACCAATCCACAT	1500
Db	1441	AACGGTATCGATTTCGCAACCTACCGTGATGTTGTACTCAAGCGTTTTACCAATCCACAT	1500
Qy	1501	ATTCAAGSACCAACCAACGCGCTCGCTCGGGATGGTTCTCGAAATTTCCGGCGGATGATT	1560
Db	1501	ATTCAAGSACCAACCAACGCGCTCGCTCGGGATGGTTCTCGAAATTTCCGGCGGATGATT	1560
Qy	1561	GCCCCACACATGCGAGAGTGCTACCAAGCAGAGCGGTTCCGCCGAATGCCACGCCCATGTTA	1620
Db	1561	GCCCCACACATGCGAGAGTGCTACCAAGCAGAGCGGTTCCGCCGAATGCCACGCCCATGTTA	1620
Qy	1621	CTGCGACTGTTTTACGTATTTCATGGAGCAGTGGCATCACGGCAAACTCGCCCTATGAATAT	1680
Db	1621	CTGCGACTGTTTTACGTATTTCATGGAGCAGTGGCATCACGGCAAACTCGCCCTATGAATAT	1680
Qy	1681	CAGATGGCATCTTGATGACACAGCTGTCATGCGAATGTTACAGTCGCGCGATCCCGTC	1740
Db	1681	CAGATGGCATCTTGATGACACAGCTGTCATGCGAATGTTACAGTCGCGCGATCCCGTC	1740
Qy	1741	GCTGTTTTATGCCAGTGATAAAGCGCTGTTTGGCGATTTTAAACGCAAGTGAAGATTTTGGC	1800
Db	1741	GCTGTTTTATGCCAGTGATAAAGCGCTGTTTGGCGATTTTAAACGCAAGTGAAGATTTTGGC	1800
Qy	1801	GCGTGTTGCGGAAAAATTCGCTACGCTCTACGCGCTAAATTAACATAA	1848
Db	1801	GCGTGTTGCGGAAAAATTCGCTACGCTCTACGCGCTAAATTAACATAA	1848

RESULT 2

```

US-09-738-626-126
; Sequence 126,, Application US/09738626
; Publication No. US20020197603A1
;
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOHO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDE
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/73738626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/3774
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/1591
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/2809
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 126
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-126

```

Query Match	3.7%	Score
Best Local Similarity	48.6%	Pred.
Matches	249	Conservative
0	Matches	0
Qy	1042	GATGTCATCGACTGGCTGAGTACAAATAC

Db	670	GAGTCGGCGAATGGGTGGAAAAACAACGTCGGCTTCCCAACTCCATCCATGGTGGACCGGCATC	729
QY	1102	ACGCCCTCGTCGGCAGCAGAACTTCCGGCAGCGATCAAGGCTCAAGCGGTATTGCCGAT	1161
Db	730	ACCCCTGAACACCGGACGCGACCGCGATGACATCAAGG--AAATCGGCTACATCGAT	786
QY	1162	AAAGCGCCGTAATGGCGGAACCTTTATCCAGTGGGTCTGTGGAAGATAAATTTCCGTGAT	1221
Db	787	GCGTGGCCATGGTTCTGAAGATTTTACCCAAATGGGTCTCGAGGATGCGTTTCCACCAG	846
QY	1222	GTCCGTCGGCAGCTGGAGAAGTGGTGTGCGAACTGTGTGGCGTCCGGTAATCCCCATATGAA	1281
Db	847	GGCGCCCGCGTACGAGAGGTTGGCGTGCAGTCTGCCGACGTGGAGCCCTTATGAA	906
QY	1282	GAGCGAAGATTGCGATTCCTTAAGTCTTTACACAGTTGTCATCGGCTGGCGAGGTACGTTA	1341
Db	907	TTAATGAAGTGGCGCTGCTCAAGCGCTCCACAGGACTTTTGTCTACTTCGGGCACATTG	966
QY	1342	ATCGGTCAAAAATATATCCACGAAGACAATACCGCATTTTATCTATCAGATTGCCGAC	1401
Db	967	GCTGGCCACCACATGGTCTCAGGAGTCATGGCGATACCCGCTTCCAGGATTTTCTCTGT	1026
QY	1402	CGTACGTGACGGAAGATGTCATTCCTTGGT---GGCGAATAACGGTATCGATTGGCA	1458
Db	1027	GCTTACATGAGGCGGAAGCACCCCTACCTCAAGAACTTCCAGGTGTGATCTAGAT	1086
QY	1459	ACCTACCGGTGATGTTGTACTCAAGCGTTTACCAATCCACATATTCAGGACACCAACAA	1518
Db	1087	GCTTATCGAGCCNACTATCGCGCGATTCGGCAACGCCGAGTCAAGACACCGTATCCG	1146
QY	1519	CGGTCGCTCGGATGGTTTTCGAAAAATGCC	1550
Db	1147	CGCTGTGTGGGAATCCTCCGACCGCATGCC	1178

```

RESULT 3
US-10-184-644-402
; Sequence 402, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-402

```

	Query Match	1.9%;	Score 36;	DB 9;	Length 653;
	Best Local Similarity	7.1%;	pred. No. 0.15;		
	Matches 38;	Conservative 159;	Mismatches 336;	Indels 0;	Gaps 0;
QY	603	TATTCGTAAATGCTGCTGAACATGTCGTACAGGCACTCAGTCACAGAAAGTCGCTATGT	662		
	:	:	:	:	:
	:	:	:	:	:
DB	57	KRYDNIETIOKFISKKADLLFALSFKSDPATSEINDESDHYVIMPLPLEQFMEIPSMDDR	116		
	:	:	:	:	:
	:	:	:	:	:
QY	663	GCTGGAAACCGCTCAGGCCGGAAGGGTAAAGCGAATATGACAGAGATCACCTCAAATTCAGAA	722		

Db	117	ELFFRIERGDIVIGRISSIRFEGFWVLICGSGIMKDIAHLEITALCPURDVP	PSNH 176
Qy	723	GTTGATACCGTGGCAGCGAGATTACAACCGCTGATTCTGAAGGGGCAGATCCG	AAGAC 782
Db	177	GDPLSYQTGDIIRAGIKDIDRYHEKLAIVLSYSSLPHLSGIKLGIVISSEELPY	YRS 236
Qy	783	AAAAGTGATTGCTTTTCCCGTCACCGAAGCGGGGTACTACCTGAATACCACT	CAACAAC 842
Db	237	VELNSNLSLEYENVMGSSLGFWPGVVEFLLEKLGIDESNPPLMRGLGSKNF	SEDDFAS 296
Qy	843	GGAAGTTTACAATCTTGATTTAGCGGCAGCATCTTAAAGGGGGATGCAAAACA	ANTTTACGG 902
Db	297	ALRKKOSAWLCKVIGDYDFKVRHVDAMNEYKALEIDKQNVLEALVARGALYAT	KGS 356
Qy	903	TGTTATTACCCGTATCTCTGAAGCGCGTATGCAATAACCGCGACCACTAAC	CCCTGCT 962
Db	357	LNKAIEDFELALENCPTHRNARKYLCQTLVERGGOLEEBEEXPLNABSYK	KKALDELTEFK 416
Qy	963	GAATTGGGATAAGTCGCGCATTAATGGTGAACGTTTCCATGATGCCGTGGT	TGATTCT 1022
Db	417	DAEDAQLKHYYMOKSLELRKQAEKEEKQTKKTETSAEKLRLKEEKLKKRKR	ST 476
Qy	1023	CCAGCTAACTGGCAACAGAGATGTCATCGACTGGCTGAGTACAATACCACT	TGCCCGGAA 1082
Db	477	SSSVSSADESVSSSSSSGSHKHKHKNRSESSRRSHSSRASNNIDQNR	DEC 536
Qy	1083	TACCATGTTGACCGATTACGCTCGTCGCGCAGCAGAACTTCGGCACCGGA	1135
Db	537	YVPANTASFLNHHOEVKLLGKDRLOYEKTQIKEDRCPLSSSSLEIPD	589

```

RESULT 4
US-10-184-644-644-c
; Sequence 346, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2003-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-346

```

	Query Match	1.8%;	Score 33.2;	DB 9;	Length 671;
	Best Local Similarity	7.3%;	Pred. No. 1.4;		
	Matches	17;	Conservative 100;	Mismatches 117;	Indels 0; Caps 0;
Qy	1443	CGGTATCGATTGGCAACCTTACCCTGATGTTGTACTCAAGCTTTTACCAATCCACAATAT	1502		
		: : : : : : : : : : : : : : : : : : :			
Dd	367	CSCY.T.CYHTY.YYY.M..M..Y.YV.....YYYYT...YYVT.HYYMW.S.YHB.H	308		
Qy	1503	TCAGGACACCAACGCGTGCTGCGGATGGTTTCTCGAANAATTCGGCGGATGATGC	1562		
		: : : : : : : : : : : : : : : : : : : :			
Dd	307	SHSHSSSSS..Y..M.MVCY.M...M..T.MVCY..MMBSBHSHSSSSSSSTSYTYKTB	248		

```
QY 1563 CCCACACTCGGAGAGTGCTTACCAGGCGGTTCGCCGAATGCCACCCCATGTTACC 1622
DB 247 ...MTCSHSMTHSHSTS.TMMMYCC...CY.TYBTMM...A.H.HSAM.S.SSS.SN.188
QY 1623 TGCACGTGTTTACGTATTTCATGCGAGCAGTGCGCATCACGGCAAACTGCCCTATCA 1676
DB 187 ..S.SBST.H.HSSTWYTYMSBSKM.T.AMYM.CSNHSMHSHSHS.KYHSTTA 134

RESULT 5
US-09-735-705-134
; Sequence 134, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy R.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C14
; CURRENT APPLICATION NUMBER: US/09/735,705
; CURRENT FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 419
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 134
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc_feature
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = A,T,C or G
US-09-735-705-134

Query Match 1.8%; Score 33.2; DB 10; Length 4797;
Best Local Similarity 51.3%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1074 TTGCCCGAATACCATGTTGACCGCATTAGCCCTCGTCCGGCAGCAGAACTTCCGGCAGC 1133
DB 3160 TGGCTGGAAGATCCTCGGGGTGGGCTTGCGGCTCACACACCTGTACTGTTAGT 3219
QY 1134 GATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCA 1193
DB 3220 GACCAAGCATCTTGGGGGTGGCGCTGAGTGGCAGGGGACAGGAGTCACTTTGTTTCG 3279

RESULT 6
US-09-850-716A-134
; Sequence 134, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
; APPLICANT: Kalos, Michael D.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07
```

```
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 134
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc_feature
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = A,T,C or G
US-09-850-716A-134

Query Match 1.8%; Score 33.2; DB 10; Length 4797;
Best Local Similarity 51.3%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1074 TTGCCCGAATACCATGTTGACCGCATTAGCCCTCGTCCGGCAGCAGAACTTCCGGCAGC 1133
DB 3160 TGGCTGGAAGATCCTCGGGGTGGGCTTGCGGCTCACACACCTGTACTGTTAGT 3219
QY 1134 GATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCA 1193
DB 3220 GACCAAGCATCTTGGGGGTGGCGCTGAGTGGCAGGGGACAGGAGTCACTTTGTTTCG 3279

RESULT 7
US-09-880-107-3939
; Sequence 3939, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3939
; LENGTH: 4797
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: unsure
; LOCATION: (1)...(4797)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3939

Query Match 1.8%; Score 33.2; DB 10; Length 4797;
Best Local Similarity 51.3%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1074 TTGCCCGAATACCATGTTGACCGCATTAGCCCTCGTCCGGCAGCAGAACTTCCGGCAGC 1133
DB 3160 TGGCTGGAAGATCCTCGGGGTGGGCTTGCGGCTCACACACCTGTACTGTTAGT 3219
QY 1134 GATCAAGGCTCAAAACGGGTATTGCCGATAAAGCGCGGTAATGGCGAAACCTTTATCCA 1193
DB 3220 GACCAAGCATCTTGGGGGTGGCGCTGAGTGGCAGGGGACAGGAGTCACTTTGTTTCG 3279

QY 1194 GTGGGTCGTGGAAGATAATTTCCGTGATGT 1223
```

Db 3280 TGGGAGGCTTAATCTAGATATCGACTTGT 3309

RESULT 8

US-09-897-778-134
; Sequence 134, Application US/09897778
; Patent No. US20020147143A1

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.455C16

CURRENT APPLICATION NUMBER: US/09/897,778

CURRENT FILING DATE: 2001-06-28

NUMBER OF SEQ ID NOS: 467

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 134

LENGTH: 4797

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: 135, 501, 4421, 4467, 4468, 4698

OTHER INFORMATION: n = A,T,C or G

US-09-897-778-134

Query Match 1.8%; Score 33.2; DB 10; Length 4797;

Best Local Similarity 51.3%; Pred. No. 5;

Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 1074 TTGCCCGAATACCATGTTGACCGCATTCACCGCTCGTCCGGCAGCAGAACTTCGGGCACG 1133

Db 3160 TGGCTGGAAGATCTCGCGGTGGGCTTGGGCTCACACACCTGTAGCATTACTGGTAG 3219

Qy 1134 GATCAAGCTCAACGGGTATTGCGGATAAAGCCGCGGTAAATGGCGAAACCTTTATCCA 1193

Db 3220 GACCAAGCATCTTGGGGGGTGGCGCTGAGTGGCGGGACAGGAGTCACTTTGTTTCG 3279

Qy 1194 GTGGCTGTGGAAGATAATTTCCGTGATGT 1223

Db 3280 TGGGAGGCTTAATCTAGATATCGACTTGT 3309

RESULT 9

US-10-184-644-310/c

; Sequence 310, Application US/10184644

; Publication No. US20030044930A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: ACIDS ENCODING THE SAME

FILE REFERENCE: P3430R1C227

CURRENT APPLICATION NUMBER: US/10/184,644

CURRENT FILING DATE: 2002-06-28

Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-184-644-310

Query Match

Best Local Similarity 17.1%; Score 32.8; DB 9; Length 596;

Matches 68; Conservative 108; Mismatches 221; Indels 0; Gaps 0;

Qy 1284 GCGGAAGATTCGCATTCTTAACCTTTCACACAGTTGCGCTCGGCGAGGTACCTTTAAT 1343

Db 471 SSCSNATANTC SSAASS.SNABAS.CSSBASS.SNATAST.SSAASHSNATASTCSSAA 412

Qy 1344 CGGTCAAAATATATCCACGAAGACAAATGACCGATTTTATCTATCAGATTGCCGACG 1403

Db 411 SS.SNATASBCSSAASS.SNATASTCSSBASS.SNATANTCSSAASS.SNATANTCSS.A 352

Qy 1404 CTACGTGACGAAGATTCATTCTTGTGGCGGATTAACGGTATCGATTTGCCAACCTA 1463

Db 351 SS.SNABAS.CSSBASS.SNATACTCSSAASHSNATASTCSSAASS.SNATASTCSSAA 292

Qy 1464 CCGTGATCTTGTACTCAAGCGTTTACCAATCCACATATTCAGGACACCAACGCGT 1523

Db 291 SS.SNATASTCSSBASSSNATASTCSSAASS.SNATACTCNAAYS.SNATASTCSSAA 232

Qy 1524 CGCTCGGATGTTCTCGAAAATTCGCGGATGATTCGCCACACACGAGAGTGCTA 1583

Db 231 SS.SNATASTYSSBASS.SNATASTCSSAASS.SNATASTCSSAASHSNATASTCSS.A 172

Qy 1584 CCAGCGAGCGTTCGCCCGAATGCCACGCCCATGTTTACCTGCACGTGTTTACGTATTCA 1643

Db 171 SS.SNATASTYSSBASS.SNATASTCSSABSSCSNABASTCSS.ASS.SNATASTCSSAA 112

Qy 1644 GGAGCGTGGCATCAGCGCAAACTCCCTATGCAATAT 1680

Db 111 SS.SNAT.S.CSSTAS..SNATAS.CSSAAD..SNAB 75

RESULT 10

US-09-815-242-4050/c

; Sequence 4050, Application US/09815242

; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: Prokaryotes

FILE REFERENCE: ELITRA.011a

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4050
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4050

Query Match      1.8%; Score 32.8; DB 10; Length 615;
Best Local Similarity 51.4%; Pred. No. 1.8;
Matches 76; Conservative 0; Mismatches 0; Gaps 0;

QY 1645 GACCAATGCGATCACCGCAACTGCCCTATGATATACAGATGGCATCCTTGTATGACCA 1704
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 GGCATGCGCGATGATGGGATGCTCCGATGAAGTCATGATGGAGTGTCCATGACGG 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1705 GCTGTCTCATGCAATGTTACAGTCTCGCATCCGTCGCTGTTTATGCCAGTGATAACGC 1764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 CGAGCGCATGCACTGAGGATGATGCCGATGAGCGATGGAGATGGAGATGTTCAAGAA 95
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1765 CTGTTTGGCGATTTAACCGAAGCTGAAG 1792
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 CTGCTGCATGCTGATGGCGAAGATGATG 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-10-184-644-258/c
; Sequence 258, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RLC227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper of Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 258
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-258

Query Match      1.8%; Score 32.8; DB 9; Length 919;
Best Local Similarity 8.7%; Pred. No. 2.3;
Matches 43; Conservative 137; Mismatches 316; Indels 0; Gaps 0;

QY 924 AGCGCGTATGGCAATATACCGCGGACCACTAACCCCTGCTGTAATTCGGATAAGTCGCGCA 983
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 512 ASH..BAHNKWTNSN A.CMS.....SM.S.HANCSA.TC.TH..CNN.T.HSBR.DSCCA 453
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 984 TAATGGTGAACGTTTCCATGATGGCTGTTGATTTCTCCAGCTAACTAGCGGAACAGGA 1043
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 452 .MSK..BT.HTTYC.T..DB.TCS.MB.H.GSSTANH.CHA..B.SCH..SD..C.B.. 393
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1044 TGTCATCGACTGCTGAGTACAAATACCACTTCCCGCAATACCATGGTTGACCGCATAC 1103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 TRM.CSG.SACC..RA..CTK.ANY.HSSM.....MMB.ATASH.DKCBWSCN.BA.... 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1104 GCCTCTCGCGGACGACGAACTTCCCGCAGCGATCAAGGCTCAACGCGGTATTGCCGATAA 1163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 DMTT.NKYN.YHMCCKSCSMH.B.GB.Y.S.M..S.B.....ABK..ANM.H.SNS.B.W 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 1164 AGCGCGGTAAATGGCGCAAAACCTTTATCCAGTGGTGGTGGGAAGATAAATTCCTGATGT 1223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 ASY.NGM.N..S.T..NDAM.NG..BBSH.S.K.K.STM.A.BMH....GHMCR.MAASH 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1224 CGTCCGCGCACTGGAGAAGTCGGTCTCGAACTGGTGGTGGTAAATCCCTTATGAAGA 1283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 .YGTYS.GSCC.GMRBYNYCS.CTSGYAT..MMSMTYR...H.NR.H.BCWY..DTW.DB.. 153
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1284 GCGAAGATTCGATTCCTTAACCTCTTCACACAGTTCATCGCTGGCGAGGTACGTTAAT 1343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 .MC..CR.N.MMC...H.A.D.R.CM.CG.A..MAR..HYC...A..TB.BHTDMDN.DM.. 93
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1344 CGGTCAAAATATATCCAGAAAGCACAAATGACCGCATTTTATCATCAGATTGCCGACCG 1403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 YMR..N.MWN.....SBNM...YM.AT...ASTAABKH.....M.H..BS.H.B.H 33
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1404 CTACGTGACGGAAGAT 1419
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 ..CNNN.M..SANS.D 17
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-10-114-170-264/c
; Sequence 264, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Burland, Valerie
; APPLICANT: Perna, Nicole T.
; APPLICANT: Plunkett, Guy
; APPLICANT: Welch, Rod
; TITLE OF INVENTION: NO. US20030023075A1 Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 264:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1560 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 264:
US-10-114-170-264

Query Match      1.8%; Score 32.4; DB 9; Length 1560;
Best Local Similarity 62.2%; Pred. No. 4.4;
```


Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 765 AGGGCAGATCGAAGACAAAGTATTGCTTTACCGTACCGAAGCGGGTACTACCT 824
 Db 225 AGGCACAGTCGCTAGACAGAGTCTTCCCTTACCGTACCGAAGCGGATACCTCCC 166

Qy 825 GAATACACAGTCACAACTGGAA 846

Db 165 CGGATCCATCGTCATTCGGTAA 144

RESULT 13

US-09-815-242-7778/c

; Sequence 7778, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7778

; LENGTH: 2436

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(2436)

US-09-815-242-7778

Query Match 1.8%; Score 32.4; DB 10; Length 2436;

Best Local Similarity 53.1%; Pred. NO. 5.9;

Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 553 CACCGTTTTCAGGTGATGGCGGATAAACGCTGGAGCATTTGTCGGGCAATATTCGTAAT 512

Db 1857 CGCCTTCGCGCGATACGCGGGATAGCTCTCGACCTTCTTGGCGGACGATCATCAT 1798

Qy 613 GATGCTGAACATGCTGACAGGCACTAGTCACAGAAAGTGCCTATGTGCTGGAACC 672

Db 1797 GTCGCGGATTCGTCGACCAACCACCATGCTGTCGAGGCTGCTCACTGCGGCGGCTC 1738

Qy 673 GTCAGCCCGG 682

Db 1737 GTCGTCGCGG 1728

Qy 673 GTCAGCCCGG 682

Db 1737 GTCGTCGCGG 1728

RESULT 14

US-10-114-170-62/c

; Sequence 62, Application US/10114170

; Publication No. US20030023075A1

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; Burland, Nicole T.

; Perna, Nicole T.

; Plunkett, Guy

; Welch, Rod

; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/114,170

; FILING DATE: 01-Apr-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/453,702

; FILING DATE: 03-DEC-1999

; APPLICATION NUMBER: 60/110,955

; FILING DATE: 04-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.

; REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 960296.95017

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 251-5000

; TELEFAX: (608) 251-9166

; INFORMATION FOR SEQ ID NO: 62:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 61663

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; SEQUENCE DESCRIPTION: SEQ ID NO: 62:

US-10-114-170-62

Query Match 1.8%; Score 32.4; DB 9; Length 61663;

Best Local Similarity 62.2%; Pred. NO. 49;

Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 285 AGGGCAGATCCGAGACAAAAGTATGCTTTACCGTACCGAAGCGGGTACTACCT 344

Db 38189 AGGCACAGTCGCTAGACAGAGTCTTCCCTTACCGTACCGAAGCGGATACCTCCC 38130

Qy 345 GAATACACAGTCACAACTGGAA 366

Db 38129 CGGATCCATCGTCATTCGGTAA 38108

RESULT 15

US-09-815-242-9961

; Sequence 9961, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:18:35 ; Search time 2836.31 Seconds
(without alignments)
10552.182 Million cell updates/sec

Title: US-09-802-208B-1

Perfect score: 1848

Sequence: 1 atgaacgacaaattacatg.....tctacgcgttaactaa 1848

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	90.8	4.9	801	17	BH395246 AG-ND-144
C 2	88.2	4.8	761	17	A2139030 SP.0178.B
C 3	81.6	4.4	506	12	BG608411
C 4	62.4	3.4	799	17	CNS01QW5
C 5	39.2	2.1	244	17	A2578838
C 6	37.8	2.0	391	17	AQ758091 HS_5552.A

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	37.6	2.0	632	13	BG922047
C 8	37.2	2.0	624	17	A2115205
C 9	36.6	2.0	787	17	CNS010B7
C 10	36.4	2.0	550	12	BG268636
C 11	36.4	2.0	581	10	AW165487
C 12	36.4	2.0	619	10	AW171987
C 13	36.4	2.0	771	17	AQ500296
C 14	36.4	2.0	791	11	AY110366
C 15	36	1.9	533	17	A2163357
C 16	35.8	1.9	684	9	AA570826
C 17	35.6	1.9	404	12	BG724574
C 18	35.6	1.9	415	13	BM321987
C 19	35.6	1.9	419	9	AI757652
C 20	35.6	1.9	454	13	BM321994
C 21	35.6	1.9	483	12	BG561744
C 22	35.6	1.9	484	12	BG561164
C 23	35.6	1.9	485	10	BE028322
C 24	35.6	1.9	817	12	BG418689
C 25	35.4	1.9	334	12	BF420883
C 26	35.2	1.9	423	12	BE848253
C 27	35.2	1.9	677	17	A2568609
C 28	35.2	1.9	687	17	A2568679
C 29	35.2	1.9	716	14	BQ482891
C 30	35.2	1.9	975	17	CNS01IP2
C 31	35	1.9	486	10	AV634532
C 32	35	1.9	515	14	BQ809588
C 33	35	1.9	527	17	A2219883
C 34	35	1.9	718	14	BQ822949
C 35	35	1.9	782	9	AI068997
C 36	35	1.9	891	17	CNS03EWK
C 37	34.8	1.9	121	17	A2579022
C 38	34.8	1.9	488	14	BQ163971
C 39	34.8	1.9	535	14	BQ163790
C 40	34.8	1.9	541	13	BM660708
C 41	34.8	1.9	700	9	AL628755
C 42	34.8	1.9	1201	17	CNS01652
C 43	34.6	1.9	481	13	BM370033
C 44	34.6	1.9	649	10	BE049425
C 45	34.6	1.9	1101	17	CNS00LXJ

ALIGNMENTS

RESULT 1
BH395246/c
LOCUS
DEFINITION
AG-ND-144F15.TR ND-TAM Anopheles gambiae genomic clone AG-ND-144F15
DNA sequence.
BH395246
BH395246.1 GI:17341387
GSS
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
1 (bases 1 to 801)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-144F15.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae

BH395246 801 bp DNA linear GSS 11-DEC-2001
AG-ND-144F15.TR ND-TAM Anopheles gambiae genomic clone AG-ND-144F15
DNA sequence.
BH395246
BH395246.1 GI:17341387
GSS
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
1 (bases 1 to 801)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-144F15.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES

source
1. .801
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-144F15"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 210 a 162 c 156 g 273 t
ORIGIN

Query Match 4.9%; Score 90.8; DB 17; Length 801;
Best Local Similarity 50.8%; Pred. No. 1.4e-16;
Matches 243; Conservative 0; Mismatches 1232; Indels 3; Gaps 1;
QY 892 ACAATTTACGGTCTATTACCGCTATCTCGAAGCGCGTATGGCAAATACACGCGGACCA 951
DB 478 ACTGTTTCGGTTCATTCGGGAGGCTCTCGTCTCGGAAATCCAAAGCAATGCGAGT 419
QY 952 CTAAACCTGCTGAATTCGGTAACAGTGGCCCAATATGGTGAACGTTTCCCATGATGCGCTG 1011
DB 418 ATTACCAATTCCTCTGTGACAACTTCAACATAACGGCAATACTGCCCAAAATGCAATTT 359
QY 1012 GTTGAGTTTCTCAGCTAACTGCAACAGGATGTCATCGACTGGCTGAGTACAATACC 1071
DB 358 TCCTCAATTTATGCA---AGCTCAGGATAAAGATTTGGCAGAATGGTTTCAGCAAAATGTA 302
QY 1072 ACTTGGCCGAATACCATGTTGACCGCATTTACGCCCTCGTCCGGCAGCAGAACTTCGGGCA 1131
DB 301 ACCTTCCCAACAGTATGGTAGACAGATAACACCTGTACTACATAGAGATGTAATA 242
QY 1132 CGGATCAAGCTCAACAGGGTATTCGGATTAAGAGCCCGGTAAATGGCGGAAACCTTTATC 1191
DB 241 AGGCTAAACGAAAGTGGTATCGCAGATAAGGCTCCGGTATATTGCGAAGATTTTACC 182
QY 1192 CAGTGGTGTGGAAGATAATTTCCGTGATGTCGGTCCGGCAGTGGAGAGTTCGGTGC 1251
DB 181 CAATGGGTAAAGACAAATTTTATTCGGGAAGACCTGCTCGGGAAGAGTGGGGTT 122
QY 1252 GAATGGTGGCGTCGCTAATCCCTATGAAGAGCGGAAAGATTCGCATTTCTTAACCTTCA 1311
DB 121 ACTTTCACAAAGATGTAAGCGGCTTACGAAATATGAAGTGAAGTCTGCTAATGCTTCT 62
QY 1312 CACAGTTGCATCCCTGGGAGGTAGCTTAATCGGTCAAAATATATCCACGAAAGCA 1369
DB 61 CATACCCCTTCTATCTCTATCTCTTCATTTTATAGCCGGATACCGCAAAAGTAGACGAAGCA 4

RESULT 2

AZ139030/c
LOCUS
SP_0178_B2_F03_T7A Strongylocentrotus purpuratus, purple sea urchin
clone plate-178 Col-6 Row-L, DNA sequence.
AZ139030
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoida; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 761)
Cameron, R.A., Mahaitas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray
, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and

TITLE
JOURNAL
MEDLINE
COMMENT
Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 178 row: L column: 6
Seq primer: T7
Class: BAC ends
High quality sequence stop: 761.

FEATURES

source
1. .761
Location/Qualifiers
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate-178 Col-6 Row-L"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli
DH10B"
BASE COUNT 173 a 204 c 184 g 200 t
ORIGIN

Query Match 4.8%; Score 88.2; DB 17; Length 761;
Best Local Similarity 50.3%; Pred. No. 8.4e-16;
Matches 216; Conservative 0; Mismatches 213; Indels 0; Gaps 0;
QY 1072 ACTTGGCCGAATACCATGTTGACCGCATTTACGCCCTCGTCCGGCAGCAGAACTTCGGGCA 1131
DB 459 ACTTTCCCATCAACAATGGTGACCGTATTGTGCCCGCAGTGACAGAGGATACGCTGGCG 400
QY 1132 CGGATCAAGCTCAACGGGTATTCGGATTAAGCGCCGGTAAATGGCGGAAACCTTTATC 1191
DB 399 AAAATCGAAACAACTTACCCTGTCGCGGATTCCTCGCGGCGTTCCTGTGAACCTTTCCCG 340
QY 1192 CAGTGGTGTGGAAGATAATTTCCGTGATGTCGGTCCGGCAGTGGAGAGTTCGGTGC 1251
DB 339 CAGTGGGTAAAGAAATTAATTTGTTGCCGAGCTCCGGAAATGGGAAAGCGGAGCC 280
QY 1252 GAATGGTGGCGTCGCTAATCCCTATGAAGAGCGGAAAGATTCGCATTTCTTAACCTTCA 1311
DB 279 GAATGGTGGTGGTGTGCTGCCCTTATGAAGAGATGAAGTTGGCATGCTCAACGGCAGT 220
QY 1312 CACAGTTGCATCCCTGGGAGGTAGCTTAATCGGTCAAAATATATCCACGAAAGCA 1371
DB 219 CATTCATCTCTGGCGTATCTGGGGTATCTTGCAGGATATCAGCACATTAATGACTGTATG 160
QY 1372 ATGACCGATTTTATCTATCAGATTGGCGACCGCTAGCTGACGGAAGATGTCATTCCTTGC 1431
DB 159 GAAGATGAACATTAATCGTTATGCGGCGTATGGCTTGTGAGGAAACGCGCGACG 100
QY 1432 TTGGGGGATAACGGTATCGATTTGCCAACTACCGCTGATGTTGTACTCAAGCGTTTACC 1491
DB 99 TTGAAAGTGCAGGCGCTTGATTTGCAAGATTACGCTCAACCGATTAAATGCACGCTATAGC 40
QY 1492 AATCCACAT 1500
DB 39 ACCCGCGT 31

RESULT 3

BG608411
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
BG608411
296895 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
BG608411
BG608411.1 GI:13658389
EST.
p1g.

ORGANISM	Sus scrofa
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. 1 (bases 1 to 506)
AUTHORS	Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keeler,J.W.
TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL	Unpublished (2000)
COMMENT	Contact: Smith rpl USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
FEATURES	PCR Primers FORWARD: AGGAACACGCTATGACCAT BACKWARD: GTTTTCCAGTCACGAC Plate: 1 row: A column: 22 Seq primer: ATTAGTGACACTATAG. Location/Qualifiers 1..506 /organism="Sus scrofa" /db_xref="taxon:9823" /clone_lib="MARC IPiG" /tissue_type="pooled" /lab_host="PH108" /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI; Library made from pooled tissue from day 11, 13, 15, 20.

BASE COUNT	130 a	118 c	138 g	120 t	
ORIGIN					
Query Match	4.4%	Score 81.6;	DB 12;	Length 506;	
Best Local Similarity	49.1%	Pred. No. 7	4e-14;		
Matches 216; Conservative 0;	Mismatches 224;	Indels 0;	Gaps 0;		
QY	1112	CGGCACGACAACTTCGGCGACGGATCAAGGCTCAACGGGTATTCGCCGATAAAGCGCGG	1171		
DB	1	CTGACCGCAATACTGGTGGAAATCAGCCAGCATCTGGGGGTGAATGATCCCTGCGCGA	60		
QY	1172	TAATGGCGAAACCTTTATCCAGTGGGTCTGTGAAGATAATTCCGGTGATGTCGTCGCGG	1231		
DB	61	TTAGCTGCGNACGGTTTATCCAGTGGGTGGTGGAGAGATAACTTCATCGCTGGGGGCTCTG	120		
QY	1232	CACCTGGAGAAAGTGGGTGCGAACTGGTGGCGTGGGTAAATCCCTATGAAGAGCGCGAAG	1291		
DB	121	CTTGGGAAGTTCGCGAGATGTACAATGGTGAATGATGCTCTGCCATGGGAAGAGATGAAC	180		
QY	1292	TTGCGCATCTTAACTCTTCACACAGTTGCATCCCTGGGCGAGGTACGTTAATCGGTCAA	1351		
DB	181	TGGCGATGCTTAATGGCAGGCATCTTTCTTCGCTATCTGCGGTACCTCTCAGGATTCG	240		
QY	1352	AATATATCCAGAAAGCAAAATGACCGATTTTATCTATCAGATTGCGGACCGGTACGTGA	1411		
DB	241	CCCATATCAGTGATTCGATCGAGGATCGCGCAATTTGCCATCGCGCCAGAACAATTAATGC	300		
QY	1412	CGAAAGATGTCATCTCTGTTGGGCGATAACGGTATCGATTTGCCAACCTACCGGTGATG	1471		
DB	301	TGGATGAGCAAGCGCCGACACTGCAAAATTAAGATGTGCGATTTAAACAACAATATCGGATA	360		
QY	1472	TTCTACTCAAGCGTTTTTACCACATATTCAGGACACCAACCAACCGGTCTGCTCGG	1531		
DB	361	AGTTAATTGCACGTTTTTGCTTAATCCGGCGCTGAAACATAAGACCTGCGCAATCGCGATGG	420		
QY	1532	ATGGTTTTCGAAAATTCGG	1551		
DB	421	ATGGCAGCCAGAAATTAACG	440		

RESULT 4	CNS01QW5	799 bp	DNA	linear	GSS 14-JUN-2001	
LOCUS	CNS01QW5/c					
DEFINITION	Anopheles gambiae GSS SP6 end of clone 31M20 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.					
ACCESSION	AL156645					
VERSION	AL156645.1	GI:7017564				
KEYWORDS	GSS.					
SOURCE	African malaria mosquito.					
ORGANISM	Anopheles gambiae					
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.					
REFERENCE	1 (bases 1 to 799)					
AUTHORS	Direct Submission					
TITLE	Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage					
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)					
	- Web : www.genoscope.cns.fr					
REFERENCE	2 (bases 1 to 799)					
AUTHORS	Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.					
TITLE	Direct Submission					
JOURNAL	Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France					
COMMENT	This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.					
FEATURES	Location/Qualifiers					
source	1..799					
	/organism="Anopheles gambiae"					
	/strain="PEST"					
	/db_xref="taxon:7165"					
	/clone="31M20"					
	/clone_lib="NotreDamel"					
	/note="end : SP6"					
BASE COUNT	151 a 245 c 206 g 167 t	30	others			
ORIGIN						
	Query Match	3.4%;	Score 62.4;	DB 17;	Length 799;	
	Best Local Similarity	51.4%;	Pred. No. 8.6e-08;			
	Matches 164;	Conservative	1;	Mismatches 151;	Indels 3;	Gaps 2;
QY	1051	GACTGGCTGAGTACAATACCATTGCCGAATACCATGGTTGACCCGATACCCCTCGT	1110			
Db	332	GACTGGATCGAATCCCACTGTTAGCTTCCCGTCAACGATGGTGCATGTCGTCTCGG	273			
QY	1111	CCGGCAGCAGAACTTCGCGGCACGGATCAAGGCTCAAACGGTATTGCCGATAAAGCCCG	1170			
Db	272	TC-ACCCCGACACGCTGGGATAAATTGAAGTTACCGCGTGCGCGATCCGGCGSG	214			
QY	1171	GTAATGGCGGAAACCTTTATCCAGTGGGTCGTGGAAGATAATTTCCGTGATGTCGCTCCG	1230			
Db	213	GTGNCNTGTCAANNNNNNNNNTNANNNNNNRATCAAGATAACTTTGTGGCGCGCGTCCA	154			
QY	1231	GCATCGAGAAAGTCGGTGTGCAACTGGTGGCTCGGTAAATCCCTATGAAGAGCGGAG	1290			
Db	153	GCGTGGGACAAACCGGAGCGGAGCTGGTGTGACAGCTGATCCCGTTTGAAGATGA--	96			
QY	1291	ATTCCGATCTTCTAACTCTTCACACAGTTGTCATGCCCTGGGCAGGTACGTTAATCGGTCAA	1350			
Db	95	ATGCGCATGCTTAACGGCAGCCACTMAATCTCGCGTACCTGGGCTATCTGGCGGGCTAT	36			
QY	1351	AAATATATCCAGGAAGCA	1369			
Db	35	CAGMATATTAACGAATGTA	17			

RESULT 5			
AZ578838			
LOCUS	A2578838	244 bp	DNA linear
			GSS 08-DEC-2000

DEFINITION	28a07 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium sp. NGR234 genomic clone 28a07, DNA sequence.																	
ACCESSION	A2578838																	
VERSION	A2578838.1	GI:11606521																
KEYWORDS	GSS.																	
SOURCE	Rhizobium sp. NGR234.																	
ORGANISM	Rhizobium sp. NGR234																	
	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium.																	
REFERENCE	1 (bases 1 to 244)																	
AUTHORS	Viprey, V., Rosenthal, A., Broughton, W.J. and Perret, X.																	
TITLE	Genetic snapshots of the Rhizobium species NGR234 genome																	
JOURNAL	Genome Biol.	1 (6), RESEARCH0014 (2000)																
MEDLINE	21114532																	
COMMENT	Contact: Virginie Viprey Laboratoire de Biologie Moleculaire des Plantes Superieures University of Geneva 1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland Tel: +44(0)1603450005 Fax: +44(0)1603450045 Email: virginie.viprey@bbsrc.ac.uk Class: shotgun.																	
FEATURES	Location/Qualifiers																	
source	1..244																	
	/organism="Rhizobium sp. NGR234"																	
	/strain="ANU265"																	
	/db_xref="taxon:394"																	
	/clone="28a07"																	
	/clone_lib="Shot-gun genomic library of Rhizobium strain ANU265"																	
	/note="Vector: M13; derivative strain of NGR234 cured of pNGR234a"																	
BASE COUNT	42 a	84 c	76 g	41 t	1 others													
ORIGIN																		
Query Match	2.1%	Score	39.2;	DB:17;	Length	244;												
Best Local Similarity	49.7%;	Pred.	No. 0.69;															
Matches	98;	Conservative	0;	Mismatches	99;	Indels	0;	Gaps	0;									
Qy	1045	GTCAATGACGTGGCTGACTACAAATATACCACATTCGCCGAATACCATGGTTGACCGCATTTACG	1104															
Db	48	GACCTGGCGGCGCTGGCGCAAGTATACGTGCGCCCTCCACCATGGTCGACCGGATCGTG	107															
Qy	1105	CCTCGTCCGGCAGCAGAACTTCGGGCACGATCAAGCTCAACGGGTATTGCCGATAAA	1164															
Db	108	CGGCCACACGGACACGCGACCGCGATCGACGCTTCGCTCGGCTCGAGGACGCC	167															
Qy	1165	CGCGCGGTATGGCGGAACCTTTATCCAGTGGGCTGTGGAAGATATTTCCGTGATGTC	1224															
Db	168	TGGCGCATCATGACCGCAACCTTCCGGCAATGGGTGATCGAGGACGATTTCCCGCTCGGT	227															
Qy	1225	CGTCCGGCACTGGAGAA	1241															
Db	228	CGCCCCGCGCTGGCANGA	244															
RESULT	6																	
AQ758091/c																		
LOCUS	AQ758091	391 bp	DNA	linear	GSS													

scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1128 row: A column: 13
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 391.
 Location/Qualifiers
 1..391
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1128 Col=13 Row=A"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACE3.6 vector at EcoRI sites"
 BASE COUNT 143 a 59 c 74 g 114 t 1 others
 ORIGIN
 Query Match 2.0%; Score 37.8; DB 17; Length 391;
 Best Local Similarity 50.8%; Pred. No. 2.4;
 Matches 90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 1324 GCCTGGCAGGTAGCTTAATCGGTCAAAATATATATCCACGAAAGCACAAATGACCGATTTT 1383
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 295 GCCTTGACTAGTACTTTGATTTTACCTTAAATTTGTACATTTTACCTACAAATGTTCAAGTTT 236
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 1384 ATCTATCAGATTCGCGACCGCTAGCTGACGGAAGATGTCATTCCTGCTGGCGGATAC 1443
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 235 TTCCTTAAGATAGATTCCTGTATGTGAGATGCTTGAAGTCATGGCATGAACATGCTAAGTT 176
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 1444 GGTATCGATTTGCCAACTACCGGTGATGTTGTACTCAAGCGTTTACCACATCCACAT 1500
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 175 TGGGACACATGCTCAACCTCCTTTCCAGAAAGACTGTACCAATTTATCTATATACTT 119
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 RESULT 7
 BG922047 632 bp mRNA linear EST 05-JUN-2001
 LOCUS BG922047.1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4950904 5',
 DEFINITION mRNA sequence.
 ACCESSION BG922047
 VERSION BG922047.1 GI:114302523
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 632)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-femail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

● ● ●

Location/Qualifiers
1. 624

/db_xref="taxon:10090"
 /clone="RPCI-23-460F24"
 /clone.lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Bra
 ECORI; Site_2: ECORI; Fe
 brain genomic DNA was is
 with a combination of EC
 selected DNA was cloned
 ECORI sites. The ligat
 DH10B electrocompetent
 161 c 129 g 18

Query Match	2.0%	Score 37.2;	DB 17;	Length 624;
Best Local Similarity	63.3%	Pred. No. 4.8;		

db 589 TGGGGATCTTCTCCAGATGTGTGAAAAATTGCAACAAGGGCTGCCCTATCTTGTCCAG 530

Db

CTCCGTGGCAGACAGCAAGTCATTTCAGAGG 500

RESULT 9
CNS010P7

LOCUS	787 bp	DNA	linear	GSS 26-JUL-
CNS010B7				

DEFINITION *Drosophila melanogaster* genome survey sequence T7 end of BAC BACN03M07 of DrosBAC library from *Drosophila melanogaster* (fruit fly)

ACCESION	fly), genomic survey sequence.
VERSION	AL098749
KEYWORDS	AL098749.1 GI:5610360
SOURCE	GSS.
ORGANISM	Drosophila melanogaster.
	Drosophila melanogaster.
	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidae; Drosophilidae; Drosophila.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 787)	Genoscope.		
	Direct Submission	Submitted (23-JUL-1999)	Genoscope - Centre National de Sequen-	
	BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)			
	- Web : www.genoscope.cns.fr			
	Determination of this BAC-end sequence was carried out as part			
	collaboration with the European Drosophila Genome Project (EDGP)			
	http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC			

FEATURES	Location/Qualifiers
source	1..787
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="BAC03M07"
	/clone_lib="DrosBAC"
	/plasmid="pBelobAC11"
	/note="end : T7"
BASE COUNT	52 a 114 c 73 g 19 t 529 others
ORIGIN	


```

RESULT 12
AW171987
LOCUS      AW171987      619 bp      mRNA      linear      EST 15-NOV-1999
DEFINITION 618047B01.y1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
sequence.
ACCESSION  AW171987
VERSION     AW171987.1      GI:6431783
KEYWORDS    ze mays.
SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 619)
AUTHORS     Walbot,V.
TITLE       Zea mays ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 618047 row: B column: 01.
            Location/Qualifiers
                1..619
                /organism="Zea mays"
                /cultivar="Ohio43"
                /db_xref="taxon:4577"
                /clone_lib="618 - Inbred Tassel cDNA Library"
                /tissue_type="tassel"
                /dev_stage="tassel"
                /lab_host="xL0LR"
                /note="Organ: tassel; Vector: PAD-GAL4-2.1 (Hybrizap);
                Inbred tassel library from Schmidl lab"
BASE COUNT  217 a 106 c 189 g 105 t
ORIGIN
Query Match      2.0%; Score 36.4; DB 10; Length 619;
Best Local Similarity 55.6%; Pred. No. 8.5;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY  981 CCATAATGGTGAACGTTTCCATGATGGCTGGTTGATGTTCTCCAGCTAACCTGGCAACA 1040
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   309 CCACAAGGTGCTTCTCCAGGAGGTCTGATGATGTTATCCAGTCAGCGGTAAGA 368

QY  1041 GGATGTCATCGACTGGCTGAGTACAAATACCACTTGCCTCCGATACATGTTGACCGCAT 1100
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   369 GGATATCTACACCGCTGATTTCTCTGAGCCCACTGGAGAAGATAAGATGATAAGAGCAT 428

QY  1101 TACGCC 1106
      |||
Db   429 CTGCC 434

RESULT 13
AQ500296
LOCUS      AQ500296      771 bp      DNA      linear      GSS 29-APR-1999
DEFINITION V4264 mtN-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', DNA sequence.
ACCESSION  AQ500296
VERSION     AQ500296
KEYWORDS    GSS.
SOURCE      baker's yeast.
ORGANISM    Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE   1 (bases 1 to 771)
AUTHORS     Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symonlatis,D., Jansen,R.,
            Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
            Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
            Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
            Gene Disruption
            Unpublished (1999)
            Contact: Kumar A
            Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
            Yale University
            P.O. Box 208103, New Haven, CT 06520-8103, USA
            Tel: 203 432 9949
            Fax: 203 432 6161
            Email: anuj.kumar@yale.edu
            te of mtN-3xHA/lacZ insertion.
            Seq primer: GGCCTTCTTCTTCTTGAAGTAC
            Class: transposon-tagged.
            Location/Qualifiers
                1..771
                /organism="Saccharomyces cerevisiae"
                /db_xref="taxon:4932"
                /clone_lib="mtN-3xHA/lacZ Insertion Library"
                /lab_host="E. coli"
                /note="Vector: pHS56-Sal; A yeast genomic DNA library
                (lacking mitochondrial DNA) was prepared in pHS56-Sal;
                genomic DNA was size-fractionated (DNA of roughly 2-3 kb
                in length) prior to cloning. This library was
                subsequently mutagenized with a mtN-3xHA/lacZ
                minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT  175 a 205 c 141 g 250 t
ORIGIN
Query Match      2.0%; Score 36.4; DB 17; Length 771;
Best Local Similarity 53.5%; Pred. No. 9.6;
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY  1506 GCACACCAACACGCGTCTCGGATGGTTTCTGAAAATTCGGCGCATGATTCGCC 1565
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   280 GGAATAATATCAGCGCGCGCGCGCGGGTGGCGAACAACATCCAGCATGAGATTCCC 339

QY  1566 CACACTGCGGAGAGTGTACACGAGCGGTTCGCCGAATGCCACGCCCATGTTACTGTC 1625
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   340 CGCGTGGGATGATTTCTCCAGCGCGGTGTCTCCGGAACACGATTCGGAGATCCTCA 399

QY  1626 ACTGTTTACGATTCATCGAG 1647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   400 ACTCTTTCCATTATAAAGGCG 421

RESULT 14
AY110366/c
LOCUS      AY110366      791 bp      mRNA      linear      HTC 25-MAY-2002
DEFINITION Zea mays CL67890_1 mRNA sequence.
ACCESSION  AY110366
VERSION     AY110366.1      GI:21214775
KEYWORDS    HTC.
SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 791)
AUTHORS     Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
            Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
            Unpublished (2002)
            Coe,E.C.
            Direct Submission
            Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
            Location/Qualifiers
                1..791
                /organism="Zea mays"
            JOURNAL     Unpublished (2002)
            REFERENCE   2 (bases 1 to 791)
            AUTHORS     Coe,E.C.
            TITLE       Direct Submission
            JOURNAL     Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
            Location/Qualifiers
                1..791
                /organism="Zea mays"

```

```
/db_xref="MaizeDB:632757"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT      154 a   193 c   129 g   230 t      85 others
ORIGIN

Query Match
Best Local Similarity 2.0%; Score 36.4; DB 11; Length 791;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 981 CCATAATGGTGAACGTTTCATGATGCGCTGGTTGAGTTTCTCCAGCTAACTGGCAAACA 1040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 CCACAAAGGTGCTTCTCCAGGAAGTGCTGATGATGTTATCCAGTCAGCGGGTAAAGA 426

QY 1041 GGATGTCATCGCTGCTGAGTACAAATACCACTGCCCAATACCATGTTGACCGCAT 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 GGATATCTACACCGGTGATTTCTTGAGCCCACTGGAGAAGATAAGATGATAAGAGCAT 366

QY 1101 TAGGCC 1106
    |||
Db 365 CCTGCC 360

RESULT 15
AZ163357
LOCUS
DEFINITION
SP_0073_B2.A10.T7A Strongylocentrotus purpuratus, purple sea urchin
, sperm genomic BAC library Strongylocentrotus purpuratus genomic
clone Plate=73 Col=20 Row=B, DNA sequence.
ACCESSION
VERSION AZ163357.1 GI:8316057
KEYWORDS
SOURCE
GSS.
ORGANISM
Strongylocentrotus purpuratus.
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 533)
Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 73 row: B column: 20
Seq primer: 77
Class: BAC ends
High quality sequence stop: 533.
Location/Qualifiers
1..533
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=73 Col=20 Row=B"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: Sperm; Vector: BAC3.6; BAC Clones in E-Coli
```

```
BASE COUNT      192 a   95 c   103 g   135 t      8 others
ORIGIN
DH10B"

Query Match
Best Local Similarity 1.9%; Score 36; DB 17; Length 533;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 96 TAAACGCTGGAGCATTGCTGCGGSCAATATTCGTAATGATGCTGAACATCTCCTACAGC 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 TAAGCAGGCTGGATCAATACGTGCCAATTCATCAAAATATTGCAATATTGNTAATAAA 343

QY 156 ACTCAGTGCACAGAAAGGTGCTATGTGCTGGAACCGTCAGCCCGGAAGGGTAAAGCGA 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 TAGAGCATCACAGAATGTCAACCAAGGATGGGAAGAGAAGAGAGATGCTCAAGAGA 403

QY 216 ATATGAAGAGATCACCTCAATTCAGAGTTGATACCGTGGCAGGCAGATTTCACACCGCT 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 AGATGGAGAGATGACCTAACATCATACATGGGAACCGCCGCTTGGACTANGATAGCTAGA 463

QY 276 GATTGCTGAAGGGGCGAGATCCGAAGA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 GACAGACGAGAATGGAAATGTTATGA 489

Search completed: March 21, 2003, 23:49:07
Job time : 2848.31 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:06:57 ; Search time 11422.3 Seconds
(without alignments)
10117.650 Million cell updates/sec

Title: US-09-802-208B-2
Perfect score: 3971
Sequence: 1 atcattgagcagttgttctt.....tcacgcgcgcagcatcgat 3971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3971	100.0	3971	1	AY005817 Escherich
2	3971	100.0	3971	6	AX268027 Sequence
3	1530.6	38.5	3233	1	AF045244 Klebsiell
C 4	824.8	20.8	10789	1	AE009731 Brucella
C 5	780.2	19.6	5930	1	AF045245 Klebsiell
C 6	643.2	16.2	10295	1	AE008252 Agrobacte
C 7	643.2	16.2	11437	1	AE009360 Agrobacte
C 8	468.6	11.8	851	1	KPNRBTDK M25606 K.aerogenes
C 9	369	9.3	10201	1	AE013623 Versinia
C 10	369	9.3	235050	1	AJ414158 Versinia
C 11	302.4	7.6	1923	9	BC014947 Homo sapi
C 12	300.8	7.6	1991	6	AX427495 Sequence
C 13	282.6	7.1	2485	9	AK090568 Homo sapi
C 14	226	5.7	1842	10	BC031708 BC031708 Mus muscu
C 15	218.6	5.5	11164	1	AE009374 Agrobacte
C 16	218.6	5.5	14945	1	AE008240 Agrobacte
C 17	205.8	5.2	2858	3	AY060286 Drosophil
C 18	204.8	5.2	323450	1	SME591790 AL591790 Sinorhizo
C 19	183	4.6	1947	9	AK022237 Y15896 Bacillus su
C 20	179.2	4.5	17787	1	BSY15896 AC015395 Drosophil
C 21	177.4	4.5	45258	2	AC015395 AF027868 Bacillus
C 22	177.4	4.5	87500	1	AF027868 AC023680 Drosophil
C 23	177.4	4.5	186218	3	AC023680 AE003480 Drosophil
C 24	177.4	4.5	310493	3	AE003480 AK054842 Homo sapi
C 25	172.2	4.3	1642	9	AK054842 X93081 B.subtilis
C 26	169.6	4.3	2664	1	BSBOFCEN X02448 Klebsiella
C 27	134.4	3.4	1361	1	KARBT AK001848 Homo sapi
C 28	124	3.1	1044	9	BC000610 Homo sapi
C 29	120.2	3.0	1180	9	BC000610 SC9727 Z48758 S.cerevisia
C 30	104	2.6	38813	8	SC9727 AE008908 Salmonell
C 31	93.2	2.3	20456	1	AF086781 Rhizobium
C 32	91.4	2.3	184	1	AF086781 AP002997 Mesorhizo
C 33	91	2.3	329709	1	AP002997 AX385150 Sequence
C 34	90.4	2.3	677	6	AX385150 ALI32674 Streptomy
C 35	75.8	1.9	39446	1	SC87 ALI16100 Botrytis
C 36	75.4	1.9	696	8	CNS01CX8 AJ243083 Brassica
C 37	73.8	1.9	1124	8	BNA243077 Streptoco
C 38	72.8	1.8	10127	1	AE010077 AE013686 Versinia
C 39	72.6	1.8	10592	1	AE013686 AJ414156 Versinia
C 40	72.6	1.8	22050	1	AJ414156 AE005436 Escherich
C 41	72.4	1.8	10346	1	AE005436 AP002560 Escherich
C 42	72.4	1.8	268857	1	AP002560 AP003137 Staphyloc
C 43	72.2	1.8	297850	1	AP003137 AP001515 Bacillus
C 44	72.2	1.8	303249	1	AP001515 AP003365 Staphyloc
C 45	72.2	1.8	341350	1	AP003365

ALIGNMENTS

RESULT 1
AY005817
LOCUS
DEFINITION Escherichia coli ribitol dehydrogenase (rtld), ribitol kinase (rtlk), and ribitol transporter (rbtr) genes, complete cds.
ACCESSION AY005817
VERSION AY005817.1 GI:11036624
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 3971)
AUTHORS Lafayette,P.R. and Parrott,W.A.
TITLE A non-antibiotic marker for amplification of plant transformation

JOURNAL	vectors in E. coli
REFERENCE	Plant Cell Rep. 20, 338-342 (2001)
AUTHORS	2 (bases 1 to 3971)
TITLE	LaFayette, P.R. and Parrott, W.A.
JOURNAL	Direct Submission
FEATURES	Submitted (26-Jul-2000) Crop and Soil Sciences, University of Georgia, Plant Sciences Bldg., Athens, GA 30602, USA
source	Location/Qualifiers
gene	1..3971
CDS	/organism="Escherichia coli"
	/strain="C"
	/db_xref="taxon:562"
	96..848
	/gene="rtld"
	96..848
	/gene="rtld"
	/codon_start=1
	/transl_table=11
	/product="ribitol dehydrogenase"
	/protein_id="AA01883.1"
	/db_xref="GI:11036625"
	/translation="MMHNSVSPMNTPLNGKVAATGAASGIGLOCAKTLDDAGAKVVLIDREGDKLHKIVAEIGENAYALQDLFNNQVDNMLADIIELAGSLDIFHANAGAYIGCPVARGDPQVMDVRLNININAFRCYRAVLPHMIAQRSDDIIFTSSIAGVVPVWEPIYTSFAFQAQFVHTTRQVQSQGVGRVAVLPQPVVITALLDDMPKAMEEALANGSLMQPIEVAESVLFWMYTRSKNVTVDLVILPGSVDL"
gene	859..2463
CDS	/gene="rtlk"
	859..2463
	/gene="rtlk"
	/codon_start=1
	/transl_table=11
	/product="ribitol kinase"
	/protein_id="AA01884.1"
	/db_xref="GI:11036626"
	/translation="MTIKTVIGVDVSGSVRAGIFDLNLSLLSHATEKITTRSGSRVEQSQELWQVSCIRNALTDVCAQSVAGIFGDFATCSLVLDKNGDPLPVSPEGDAKQNIIVMDHRTAEQERINATHHPVLNYYVGGKISPEMETPKILMLKENMPEIYERAGQFDLIDWTRATGLARSCTVCTCKTWLAHENRMDPDYFTIIGLAELADEDFIRIGHIVSPGTCGNGLAQAAMGLLPPTGVAVGLIDAHAGGIGTVGVGGALNNLAVVFTSGTMASTSPSFVPGVWPYTSAMVPLGLVVEGGSAAAGAIIDOLLDPHAPVEAREAKORNPVLPVWLADRIEKTQPSDAVALAKGLHVVPFELGNRAPFPADPHA RAVICGLGMRDLNLLALYIAGLCGIGYGLRQILDQAQVAGVSKNIVISGGAGQHP LVKQILADTGCPVITTOCEPVLIGSAILGAVAGNIAPSVGEAMQQTTHVDKYYTYP ERYQSLHRRRYEAYKQLOHTAKLRD"
gene	2565..3839
CDS	/gene="rbt"
	2565..3839
	/gene="rbt"
	/codon_start=1
	/transl_table=11
	/product="ribitol transporter"
	/protein_id="AA01885.1"
	/db_xref="GI:11036627"
	/translation="MSRNNKQWGLPLHLINGYIAIAVFTGDFELAFSLHYIKSLGFTPACASAFYLGAAALSAWVSGVAEITTPQKTMIGFVLMCVFHVLELVFGGQANGYILIFGIRGLAYPLFLYFTVIHNRVSRNSSALGWYVWVSVIGVAGSYIPSFPIPMGENTLWALAFCFAGGVIAITSLRHVKTPGHMNLTPREKFAELSRVITLLYNNRIFLSSIIVRIINTLSLFGFVIMPMWFDELFTTSEWLOYWAAFFFTTIFSNIFWGIYAEKMGWVRVIRWFCGLGMAASSLAFYVMPQYFGHYNWMIIPALAGTFVAAFVMAAVFALEKPKHGAALISVYNSAGMSNFLAPAIIVVLLPWFSTIGVVIAYITAYLLAFVLCAFIRVQPGFSSAPVTEKALNIS"
BASE COUNT	919 a 930 c 1020 g 1102 t
ORIGIN	
Query Match	100.0%; Score 3971; DB 1; Length 3971;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 3971; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATCGATTGAGCATTGCTTACACGGCAGCTAAATCCCGTTTCAGTGCCTGCAACGAA 60
Db	
QY	1 ATCGATTGAGCATTGCTTACACGGCAGCTAAATCCCGTTTCAGTGCCTGCAACGAA 60
Db	

QY	61	CAGAGACTATAAATTCGCCTCGTAAAGGATATATGATGAATCACTCTGTGCCCTCTA	120
Db	61		
QY	121	TGAATACTCCCTTAATGCAAAAGTTGCAGCTATCACTGGCGCTGCGTCAAGTATTGGCC	180
Db	121		
QY	181	TGCAATGTCAAAAACGCTGCTCGATGAGGAGCAAGGTAGTATTGATTGACCCGGAAG	240
Db	181		
QY	241	GGCACAACATGCACAAAGATTGCGTGAAGTAAAGCGGAAACGCGTACGCGCTCAACTCG	300
Db	241		
QY	301	ATCTCTTCAATATATCAGCAAGTCGATAAATGCTGCGGACATATTGCAACTGGCGGGTG	360
Db	301		
QY	361	GGCTGGATATTTTTCATGCCAATGCAAGCGCTTATATTGGCGGCCAGTGGCTGAAGGTG	420
Db	361		
QY	421	ATCCAGATCTGGGATCGTGTGTTAAATCTGAATATAAATGGGCGCTTTCGCTGTGTCC	480
Db	421		
QY	481	GTGCGAGTCTGCGCATATGATTGCGCAGAGTCGCGGCATATAATTTTACCAGTTCCA	540
Db	481		
QY	541	TGCGGGCGTCTGCTCGGTTATCTGGGAACCGATCTACACCGCTCCAAATTTGCCGTTTC	600
Db	541		
QY	601	AGCATTCGTACACACTACCGCGCCAGGTTTCTCAATATGGCGTGCCTGGGTGGCG	660
Db	601		
QY	661	TGCTGCGAGGACAGTAGTCACTGCCCTGCTGATCACTGGCCAAAGCAAAATGGAAG	720
Db	661		
QY	721	AAGCCCTGGCAAAATGATGACCAATGAACCGGATGAAAGTGGCGGAATCAGTATTGTTA	780
Db	721		
QY	781	TGCTGACCCGCTCGAAAAATGTCACCGTGGGAGATTTAGTGATCCTGCGTGGCAGTCTCG	840
Db	781		
QY	841	ATCTGTAAGGGCGCAATCATGACAATGAACAAACCGTTATTGGTGTGATGGGATCA	900
Db	841		
QY	901	GGCAGTGTCCGCGCCGGGATTTTGTATCTCAACGGATCTCTGCTATCCCATGCCAGAA	960
Db	901		
QY	961	AAATCAGACTACGCGGCGCAGCGGAGCGGATGGAGACAGTCCAGCAGGAGATCTGG	1020
Db	961		
QY	1021	CAGCGGCTGTTCATGATATCGTATGCGTCACTCTGCGACAGCTTTGTGCACAAAGT	1080
Db	1021		
QY	1081	GTGGCAGGATCGGTTTGTATGCCACCTGTTCTCTGGTGTACTGGATAAAACGGTGAT	1140
Db	1081		
QY	1141	CCATTGCGCTGTGAGCCCGGAGGAGATGCAAGCAAAATATCATTTGTTGGATGGATCAC	1200

Db 1141 |||||CCATTGCGCTGTCAGCCCGGAAGAGATCAAGCAAAATATCATTTGTGTGGATGCAC 1200
Qy 1201 CCGCGCACCGCAAGCAGAGGAATCAATGCCACTCACCATCCGGTGTGCTGAACCTACGTC 1260
Db 1201 CCGCGCACCGCAAGCAGAGGAATCAATGCCACTCACCATCCGGTGTGCTGAACCTACGTC 1260
Qy 1261 GGTGTTAAATTTTCGCCCTGAAATGGAAACACCGAAATTTCTGTGCTGAAAGAAATATG 1320
Db 1261 GGTGTTAAATTTTCGCCCTGAAATGGAAACACCGAAATTTCTGTGCTGAAAGAAATATG 1320
Qy 1321 CCAGAGATCTACGAACGTGCGGAGCAATTTTCGATCTGGCGGATTTTCTGACCTGGCGG 1380
Db 1321 CCAGAGATCTACGAACGTGCGGAGCAATTTTCGATCTGGCGGATTTTCTGACCTGGCGG 1380
Qy 1381 GCTACCGGTGATTTAGCCGTTTACGATGTCAGTGTACCTGTAAATGACGCTGGCTGGCA 1440
Db 1381 GCTACCGGTGATTTAGCCGTTTACGATGTCAGTGTACCTGTAAATGACGCTGGCTGGCA 1440
Qy 1441 CATGAAATTCGCTGGATCCAGATTTATTTCCGCACCATTTGGCCTTGCGAGTTAGCGAT 1500
Db 1441 CATGAAATTCGCTGGATCCAGATTTATTTCCGCACCATTTGGCCTTGCGAGTTAGCGAT 1500
Qy 1501 GAAGATTTTATTCGATTTGGTCATCATATGTTTCTCCGGAACACCTTTGTGGAAATGGT 1560
Db 1501 GAAGATTTTATTCGATTTGGTCATCATATGTTTCTCCGGAACACCTTTGTGGAAATGGT 1560
Qy 1561 TTAACAGCACAAAGCCGCGAGAGATGGATTTACTCCCGGCACACCTTGCCTGTAGGG 1620
Db 1561 TTAACAGCACAAAGCCGCGAGAGATGGATTTACTCCCGGCACACCTTGCCTGTAGGG 1620
Qy 1621 TTAATTGATGCTCAGCTGGTGGATCGGTACGGTTGGCTAGAAAGTGGAGCGCTGAAC 1680
Db 1621 TTAATTGATGCTCAGCTGGTGGATCGGTACGGTTGGCTAGAAAGTGGAGCGCTGAAC 1680
Qy 1681 AATCTCGGTATGTTTTCGGCACCTTCTTCATGCGACCATGGCATPACCACTTCTCCCTCG 1740
Db 1681 AATCTCGGTATGTTTTCGGCACCTTCTTCATGCGACCATGGCATPACCACTTCTCCCTCG 1740
Qy 1741 TTTGTACCGGTGTCCTGGGCGGTATTTACAGTGGATGGTTTCCAGGTCCTGGTTAGTT 1800
Db 1741 TTTGTACCGGTGTCCTGGGCGGTATTTACAGTGGATGGTTTCCAGGTCCTGGTTAGTT 1800
Qy 1801 GAAGCGGCAAGTGTCTCGGAGCAGCTATTTGACCACTACTTTGATTTCCATCCGGCT 1860
Db 1801 GAAGCGGCAAGTGTCTCGGAGCAGCTATTTGACCACTACTTTGATTTCCATCCGGCT 1860
Qy 1861 GTTGAAGAGCTCGGAAATGGCACAACGTGTGAATCAGCCCTCCCGCTCTGGCTTGGT 1920
Db 1861 GTTGAAGAGCTCGGAAATGGCACAACGTGTGAATCAGCCCTCCCGCTCTGGCTTGGT 1920
Qy 1921 GATCGAATCTCGAAAAACGCGCAACCATCAGATGCTGTGCGCCTCGCGAAAGGGCTA 1980
Db 1921 GATCGAATCTCGAAAAACGCGCAACCATCAGATGCTGTGCGCCTCGCGAAAGGGCTA 1980
Qy 1981 CACGTGGTCCGGAATTTCTCGGAAATCGCGCCCTTCGAGATPCTCATGCCAGAGCG 2040
Db 1981 CACGTGGTCCGGAATTTCTCGGAAATCGCGCCCTTCGAGATPCTCATGCCAGAGCG 2040
Qy 2041 GTAATTTGTGGCTGGGTATGGAGGAGATCTGGATAATTTACTCGCCTTGTATATCGCT 2100
Db 2041 GTAATTTGTGGCTGGGTATGGAGGAGATCTGGATAATTTACTCGCCTTGTATATCGCT 2100
Qy 2101 GGATTTATCGGAATTTGTTTGTGTCGCCAAATTTCTGACGCTCAACACGCGCAGGA 2160
Db 2101 GGATTTATCGGAATTTGTTTGTGTCGCCAAATTTCTGACGCTCAACACGCGCAGGA 2160
Qy 2161 GTAGTGAGTAAATATCGTTTATAGCGCGGTGCGCGGAGCATPCTACTGGTAGCAAA 2220
Db 2161 GTAGTGAGTAAATATCGTTTATAGCGCGGTGCGCGGAGCATPCTACTGGTAGCAAA 2220
Qy 2221 ATTTCTGGCAGATACCTCGCGGTATTTCCGGTCAATACCACGCAATGCTGCGAACCGGTTTA 2280
|||||

Db 2221 ATTTCTGGCAGATACCTCGCGGTATTTCCGGTCAATTAACACGCAATGCTGCGAACCGGTTTA 2280
Qy 2281 TTAGGCTCGGCATTTCTTTGGTGTCTGCGCGGAAATATTGCACTTTCTGTGTGCGCAAGCG 2340
Db 2281 TTAGGCTCGGCATTTCTTTGGTGTCTGCGCGGAAATATTGCACTTTCTGTGTGCGCAAGCG 2340
Qy 2341 ATGCAACAATTCACCCATGTGGATAAATATTATTTATTCGCAAGACGCTATCAATCTCTT 2400
Db 2341 ATGCAACAATTCACCCATGTGGATAAATATTATTTATTCGCAAGACGCTATCAATCTCTT 2400
Qy 2401 CATCATCTGCTGATATGAGCCCTATAAGCAGTTGCGAGCTACTGCAAAATTTACTCAGAGAC 2460
Db 2401 CATCATCTGCTGATATGAGCCCTATAAGCAGTTGCGAGCTACTGCAAAATTTACTCAGAGAC 2460
Qy 2461 TAAATTAACAGCGCTGACGCTGTTTTCAGGCAATCACTAATAACGACTCACTCCGGTA 2520
Db 2461 TAAATTAACAGCGCTGACGCTGTTTTCAGGCAATCACTAATAACGACTCACTCCGGTA 2520
Qy 2521 ATATCCCGAGTGCATTCATCTGCACCCCTAAACAGAGTTTTATATGTCAGAAATAATA 2580
Db 2521 ATATCCCGAGTGCATTCATCTGCACCCCTAAACAGAGTTTTATATGTCAGAAATAATA 2580
Qy 2581 AACAGTGGTTGGTTGGCACTGCACTGATATGGGATATATCGCCATCGCAGTATTTA 2640
Db 2581 AACAGTGGTTGGTTGGCACTGCACTGATATGGGATATATCGCCATCGCAGTATTTA 2640
Qy 2641 TGACTGGTGGTTTTCGAACCTCGCATTTCTTATCGCACTATATTAATGTTAGCTTCA 2700
Db 2641 TGACTGGTGGTTTTCGAACCTCGCATTTCTTATCGCACTATATTAATGTTAGCTTCA 2700
Qy 2701 CACCTGCGGAAGCCCTTTTTCGCTTTAGCCTCTAGCGCTGCGGCTGCGCTTTCCGCT 2760
Db 2701 CACCTGCGGAAGCCCTTTTTCGCTTTACGCTCTACGCTGCGGCTGCGCTTTCCGCT 2760
Qy 2761 GGGTTCTGGGTAGTAGCGGAAATCATCACGCCGCAAAACCACTAGTGTGTTTG 2820
Db 2761 GGGTTCTGGGTAGTAGCGGAAATCATCACGCCGCAAAACCACTAGTGTGTTTG 2820
Qy 2821 TCCTATGTGGCTATTCATGTTCTGTTCTGTTCTTTGGATTAGGACAGGCAAACTATG 2880
Db 2821 TCCTATGTGGCTATTCATGTTCTGTTCTGTTCTTTGGATTAGGACAGGCAAACTATG 2880
Qy 2881 GGTAAATCCTGCTGTTTACGGGATTCGTGGTTTACGCTATCCGCTATTTCTACTCAT 2940
Db 2881 GGTAAATCCTGCTGTTTACGGGATTCGTGGTTTACGCTATCCGCTATTTCTACTCAT 2940
Qy 2941 TTATCGTGGTTATTTATTAACAGTGGTAGCGAAAAATCCAGTTCTGCTCTGGGTGGT 3000
Db 2941 TTATCGTGGTTATTTATTAACAGTGGTAGCGAAAAATCCAGTTCTGCTCTGGGTGGT 3000
Qy 3001 ACTGGCGGTATATTTACGTAGGATCGGTGTTGCTGGCAGTTATATTCAGTGTTCACGA 3060
Db 3001 ACTGGCGGTATATTTACGTAGGATCGGTGTTGCTGGCAGTTATATTCAGTGTTCACGA 3060
Qy 3061 TACCGATTATGGTGAAATGGGAACCTTATGGCTGGCACTGGCGTCTGCTTTGCTGGCG 3120
Db 3061 TACCGATTATGGTGAAATGGGAACCTTATGGCTGGCACTGGCGTCTGCTTTGCTGGCG 3120
Qy 3121 GTGTCAATTCGATGATCTCTTCGCTCATGTTTAAACGCGCTGGACATATGCAATAATTA 3180
Db 3121 GTGTCAATTCGATGATCTCTTCGCTCATGTTTAAACGCGCTGGACATATGCAATAATTA 3180
Qy 3181 CTCGCCGTGAGAAGTTTCAGAAATTAAGTCGGGAGTAACTTTACTTTATACCAACCGCA 3240
Db 3181 CTCGCCGTGAGAAGTTTCAGAAATTAAGTCGGGAGTAACTTTACTTTATACCAACCGCA 3240
Qy 3241 ATATTTTCTCTCCAGTATTTGTGCGCAATTAATAACCTTATCGTTATTCGGTTTGGCG 3300
Db 3241 ATATTTTCTCTCCAGTATTTGTGCGCAATTAATAACCTTATCGTTATTCGGTTTGGCG 3300
Qy 3301 TCATTATGCCAATGATGTTTGTGGATGAACCTGGGATTAACCACTCTGAATGTTGTCAGG 3360
Db 3301 TCATTATGCCAATGATGTTTGTGGATGAACCTGGGATTAACCACTCTGAATGTTGTCAGG 3360
|||||

ORIGIN	Query Match						100.0%;	Score 3971;	DB 6;	Length 3971;
	Best Local Similarity						100.0%;	Pred. No. 0;		
	Matches 3971;						Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATCGATTGACGAGTTTGCTTTCACACGCGACGTAAATTC	CGCTTACGCTGCGTTCAGTCCGCTGCAAGCAAA	60						
Db										
Db	1	ATCGATTGACGAGTTTGCTTTCACACGCGACGTAAATTC	CGCTTACGCTGCGTTCAGTCCGCTGCAAGCAAA	60						
Qy	61	CAGAGACTATAAATTCGCCCTTGTA	AAAAAGGATTATATGATGAATCACTCTGTGCCCTCTA	120						
Db	61	CAGAGACTATAAATTCGCCCTTGTA	AAAAAGGATTATATGATGAATCACTCTGTGCCCTCTA	120						
Qy	121	TGAATATCCCCCTTAATGGCAAAAGTTGC	AGACTATCACTGGCGCTCGGTATTTGGCC	180						
Db	121	TGAATATCCCCCTTAATGGCAAAAGTTGC	AGACTATCACTGGCGCTCGGTATTTGGCC	180						
Qy	181	TGCAATGTGCAAAACGCTGCTCGATG	CAGAGCAAAAGGTAGTATTGACCGGGAAG	240						
Db	181	TGCAATGTGCAAAACGCTGCTCGATG	CAGAGCAAAAGGTAGTATTGACCGGGAAG	240						
Qy	241	CGGACAAACTGCACAGATTTGCTGCT	GAGTTAGGCGGAACCGGTACGCGCTGCACACTCG	300						
Db	241	CGGACAAACTGCACAGATTTGCTGCT	GAGTTAGGCGGAACCGGTACGCGCTGCACACTCG	300						
Qy	301	ATCTCTTCAATATATCAGCAAGTCGAT	AACATGCTGGCGGACATATATCGAACTGGCGGGTG	360						
Db	301	ATCTCTTCAATATATCAGCAAGTCGAT	AACATGCTGGCGGACATATATCGAACTGGCGGGTG	360						
Qy	361	GGCTGGATATTTTTCATG	CCAAATGCAAGGCGCTTATATTGGCGGCCAGTG	420						
Db	361	GGCTGGATATTTTTCATG	CCAAATGCAAGGCGCTTATATTGGCGGCCAGTG	420						
Qy	421	ATCCAGATGTCGGGATCGTGCTTAA	NTCTGAAATTAATCGCGGCTTTCGCTGTGTC	480						
Db	421	ATCCAGATGTCGGGATCGTGCTTAA	NTCTGAAATTAATCGCGGCTTTCGCTGTGTC	480						
Qy	481	GTGCAGTCTCGCGCATATGATTTGC	GCAGAGGTGCGGCGCATATAATTTTACCAGTTTCCA	540						
Db	481	GTGCAGTCTCGCGCATATGATTTGC	GCAGAGGTGCGGCGCATATAATTTTACCAGTTTCCA	540						
Qy	541	TCGGCGGCGTGTTTCGGTTATCTG	GGGAACCGATCTACACCGCGTCCA	600						
Db	541	TCGGCGGCGTGTTTCGGTTATCTG	GGGAACCGATCTACACCGCGTCCA	600						
Qy	601	AGGCATTCTGACACTACCGCGCG	CGCAGTTTCTCAATATGGCTGCGTGTGGTGCGG	660						
Db	601	AGGCATTCTGACACTACCGCGCG	CGCAGTTTCTCAATATGGCTGCGTGTGGTGCGG	660						
Qy	661	TGCTGCCAGGACCAATGATGAC	TGCCCTGCTTGTGATGACTGGCCAAAGCCAAATATGGGAAG	720						
Db	661	TGCTGCCAGGACCAATGATGAC	TGCCCTGCTTGTGATGACTGGCCAAAGCCAAATATGGGAAG	720						
Qy	721	AAGCCCTGGCAATGTGAGCTGAT	CGACACCGATTCAAGTGGCGAATCACTATTGTTTA	780						
Db	721	AAGCCCTGGCAATGTGAGCTGAT	CGACACCGATTCAAGTGGCGAATCACTATTGTTTA	780						
Qy	781	TGGTGACCCGCTCGAAATATGT	CACCGTGCAGAGATTTAGTGATCTGCTGCGCAGTGTGCG	840						
Db	781	TGGTGACCCGCTCGAAATATGT	CACCGTGCAGAGATTTAGTGATCTGCTGCGCAGTGTGCG	840						
Qy	841	ATCTGTAAAGGGCGAATCATG	ACAAATTAACCAACCGTTTATGTGTGTTGATGTGGGATCA	900						
Db	841	ATCTGTAAAGGGCGAATCATG	ACAAATTAACCAACCGTTTATGTGTGTTGATGTGGGATCA	900						
Qy	901	GGCAGTGTCCGCGCGGGATTTT	TGTGATCTCAACGGATCTCTCTATCCCATGCCACAGAA	960						
Db	901	GGCAGTGTCCGCGCGGGATTTT	TGTGATCTCAACGGATCTCTCTATCCCATGCCACAGAA	960						
Qy	961	AAATACAGACTACGCGGCG	CAGCGGAACCGCGGTGGAACTCCAGCCAGGAGATCTGG	1020						
Db	961	AAATACAGACTACGCGGCG	CAGCGGAACCGCGGTGGAACTCCAGCCAGGAGATCTGG	1020						

Qy	1021	CAGCGGCTGTTCATGTATTGCTAATGGCGTCACTCTGGCAGACGTTTGTGCACAAAGT	1080
Db	1021		1080
Qy	1021	CAGCGGCTGTTCATGTATTGCTAATGGCGTCACTCTGGCAGACGTTTGTGCACAAAGT	1080
Db	1021		1080
Qy	1081	GTGCAGGCATCGTTTTCATGCCACCTGTTCTCTGTGGTGTTACTTGGATAAAAACGGTGAT	1140
Db	1081		1140
Qy	1081	GTGCAGGCATCGTTTTCATGCCACCTGTTCTCTGTGGTGTTACTTGGATAAAAACGGTGAT	1140
Db	1081		1140
Qy	1141	CCATTGGCTGTACGCCGGAAGGAGATCAAAGCAAAATATCATTTGTGTGGATGGATCAC	1200
Db	1141		1200
Qy	1141	CCATTGGCTGTACGCCGGAAGGAGATCAAAGCAAAATATCATTTGTGTGGATGGATCAC	1200
Db	1141		1200
Qy	1201	CGCGCACCAACAGCAGACGAATCAATGCCACTCACCATCGGTGCTGGAACCTAGCTC	1260
Db	1201		1260
Qy	1201	CGCGCACCAACAGCAGACGAATCAATGCCACTCACCATCGGTGCTGGAACCTAGCTC	1260
Db	1201		1260
Qy	1261	GGTGGTAAAAATTTCGCTGAAATGGAACACCGAAATTTCTTGGCTGAAAGAAAATATG	1320
Db	1261		1320
Qy	1261	GGTGGTAAAAATTTCGCTGAAATGGAACACCGAAATTTCTTGGCTGAAAGAAAATATG	1320
Db	1261		1320
Qy	1321	CCAGAGATCTACCAACGTCGCGACAAATTTTTCGATCTGGCCGATTTCTTGACCTGGCGG	1380
Db	1321		1380
Qy	1321	CCAGAGATCTACCAACGTCGCGACAAATTTTTCGATCTGGCCGATTTCTTGACCTGGCGG	1380
Db	1321		1380
Qy	1381	GCTACCGGTGATTTACGCGGTTTCAGTATGCACTGTTACTCTGAATGCACTGGCTGGCA	1440
Db	1381		1440
Qy	1381	GCTACCGGTGATTTACGCGGTTTCAGTATGCACTGTTACTCTGAATGCACTGGCTGGCA	1440
Db	1381		1440
Qy	1441	CATGAANAATCGCTGGGATCCAGATTATTTCCGCACCATTTGGCTTTGCAGAGTTAGCGGAT	1500
Db	1441		1500
Qy	1441	CATGAANAATCGCTGGGATCCAGATTATTTCCGCACCATTTGGCTTTGCAGAGTTAGCGGAT	1500
Db	1441		1500
Qy	1501	GAAGATTTTATTCGCATTTGGTTCATCATATTGTTTCTCCCGGAACACCTTTGTGGAATTGGT	1560
Db	1501		1560
Qy	1501	GAAGATTTTATTCGCATTTGGTTCATCATATTGTTTCTCCCGGAACACCTTTGTGGAATTGGT	1560
Db	1501		1560
Qy	1561	TTAACAGCACAGCCGGCAGAGATGGGATTTACTCCCGGCACACCTGTGCTGTAGGG	1620
Db	1561		1620
Qy	1561	TTAACAGCACAGCCGGCAGAGATGGGATTTACTTCCCGGCACACCTGTGCTGTAGGG	1620
Db	1561		1620
Qy	1621	TTAATTGATGCTCACGCTGGTGGCATCGGTAGCGTTGGCTTAGAAGTGGAGCGCTGAAC	1680
Db	1621		1680
Qy	1621	TTAATTGATGCTCACGCTGGTGGCATCGGTAGCGTTGGCTTAGAAGTGGAGCGCTGAAC	1680
Db	1621		1680
Qy	1681	AATCTCGGTATGTTTTCGGCACCTTCTTCATGCACCATGGCATCTACACCTTCTCCCTCG	1740
Db	1681		1740
Qy	1681	AATCTCGGTATGTTTTCGGCACCTTCTTCATGCACCATGGCATCTACACCTTCTCCCTCG	1740
Db	1681		1740
Qy	1741	TTTGTACCGGTTGTTGGGGCGGTATTACAGTGGGATGGTTCAGAGTCTGTGGTTAGTT	1800
Db	1741		1800
Qy	1741	TTTGTACCGGTTGTTGGGGCGGTATTACAGTGGGATGGTTCAGAGTCTGTGGTTAGTT	1800
Db	1741		1800
Qy	1801	GAAGCGGGCAAAAGTGTCTCGGGAGCAGCTATTGACCAGCTACTTGATTTCATCCGGCT	1860
Db	1801		1860
Qy	1801	GAAGCGGGCAAAAGTGTCTCGGGAGCAGCTATTGACCAGCTACTTGATTTCATCCGGCT	1860
Db	1801		1860
Qy	1861	GTTGAAGAAGCTCGGAATGGCACAACTGTGTAATCAGCCCCCTCCCGCTTGCTGCT	1920
Db	1861		1920
Qy	1861	GTTGAAGAAGCTCGGAATGGCACAACTGTGTAATCAGCCCCCTCCCGCTTGCTGCT	1920
Db	1861		1920
Qy	1921	GATCGAATCTCTCAAAAAACGGCGAACCATCAGATGCTGTGCGCTTGGCGAAAGGCTA	1980
Db	1921		1980
Qy	1921	GATCGAATCTCTCAAAAAACGGCGAACCATCAGATGCTGTGCGCTTGGCGAAAGGCTA	1980
Db	1921		1980
Qy	1981	CACGTGTGCGGAATTTCTTCGGAATTCGCGCCCCCTTCGAGATCTCATGCCAGAGCG	2040
Db	1981		2040
Qy	1981	CACGTGTGCGGAATTTCTTCGGAATTCGCGCCCCCTTCGAGATCTCATGCCAGAGCG	2040
Db	1981		2040
Qy	2041	GTAATTTGCGCCTGGGTATGGAGCGAGATCTGGATAAATTTACTCGCCCTTGATATCGCT	2100
Db	2041		2100
Qy	2041	GTAATTTGCGCCTGGGTATGGAGCGAGATCTGGATAAATTTACTCGCCCTTGATATCGCT	2100
Db	2041		2100

Qy	2101	GGATTATGCGGAATTCGGTTATGGTCTGCCCAAAATTCGACGCTCAACACGCCAGGGA	2161
Db	2101	GGATTATGCGGAATTCGGTTATGGTCTGCCCAAAATTCGACGCTCAACACGCCAGGGA	2160
Qy	2161	GTAGTGAGTAAAAAATATCGTTATTAGCGCGGTGCCGGCAGCATCCACTGGTACGACAA	2220
Db	2161	GTAGTGAGTAAAAAATATCGTTATTAGCGCGGTGCCGGCAGCATCCACTGGTACGACAA	2220
Qy	2221	ATTCTGCGCAGATACCTTGCGGGTATTCGCGTCAATTACCACGCAATGCTGCGAAACCGGTTTTA	2280
Db	2221	ATTCTGCGCAGATACCTTGCGGGTATTCGCGTCAATTACCACGCAATGCTGCGAAACCGGTTTTA	2280
Qy	2281	TTAGGCTCGCGCAATCTCTGCTGCTGCGCGGAAATATTGCACCTTCTGTTGGCGAAAGCG	2340
Db	2281	TTAGGCTCGCGCAATCTCTGCTGCTGCGCGGAAATATTGCACCTTCTGTTGGCGAAAGCG	2340
Qy	2341	ATGCACAATTCACCCATCTGGATAAATAATTATTATCCGCAAGAACGCTATCAATCTCTT	2400
Db	2341	ATGCACAATTCACCCATCTGGATAAATAATTATTATCCGCAAGAACGCTATCAATCTCTT	2400
Qy	2401	CATCATCGTCGATATGAGCCCTATAGCAGTTTGCAGCATACTGCAAAATTACTCAGAGAC	2460
Db	2401	CATCATCGTCGATATGAGCCCTATAGCAGTTTGCAGCATACTGCAAAATTACTCAGAGAC	2460
Qy	2461	TAATTAAACCAAGCCGCTGACGCTGTTTTTCAGGCAATCACTAATAACGACTCACTCCGGA	2520
Db	2461	TAATTAAACCAAGCCGCTGACGCTGTTTTTCAGGCAATCACTAATAACGACTCACTCCGGA	2520
Qy	2521	ATATCCCGGAGTCGATTCATCTGCACCCCTAAACACGAGGTTTATATGTTCCAGAAATAATA	2580
Db	2521	ATATCCCGGAGTCGATTCATCTGCACCCCTAAACACGAGGTTTATATGTTCCAGAAATAATA	2580
Qy	2581	AACAGTGGTTGGGTTTGCCACTGCATCTCATATGGGGATATATCGCCATCCGAGTATTTA	2640
Db	2581	AACAGTGGTTGGGTTTGCCACTGCATCTCATATGGGGATATATCGCCATCCGAGTATTTA	2640
Qy	2641	TGACTGGTGATGTTTCGAACCTCGCAATTCGCAATTCCTTATCGCACTATATTAAATCGTTAGGCTTCA	2700
Db	2641	TGACTGGTGATGTTTCGAACCTCGCAATTCGCAATTCCTTATCGCACTATATTAAATCGTTAGGCTTCA	2700
Qy	2701	CACCTCGGGAAGCCTCTTTTGGCCTTTACGCTCTACGGCTGCGGCTCCGCTTTCCGCT	2760
Db	2701	CACCTCGGGAAGCCTCTTTTGGCCTTTACGCTCTACGGCTGCGGCTCCGCTTTCCGCT	2760
Qy	2761	GGGTTCTCGGGTAGTAGCGGAAATCATCACGCCGCAAAAAACCATGCTGATGGTTTG	2820
Db	2761	GGGTTCTCGGGTAGTAGCGGAAATCATCACGCCGCAAAAAACCATGCTGATGGTTTG	2820
Qy	2821	TCCTATGGTCGGTATTTCCATGTTCTGTTTCTGGTCTTTGGATTAGGACAGGCAAACTATG	2880
Db	2821	TCCTATGGTCGGTATTTCCATGTTCTGTTTCTGGTCTTTGGATTAGGACAGGCAAACTATG	2880
Qy	2881	GGTTAAATCCTGCTGTTTTACGGGATTCGTGGTTTAGCGTATCCGCTATTTCTCTACTCAT	2940
Db	2881	GGTTAAATCCTGCTGTTTTACGGGATTCGTGGTTTAGCGTATCCGCTATTTCTCTACTCAT	2940
Qy	2941	TTATCTGCTGTTATTATTCATAAGCTCGGTAGCGAAAAATTCAGTTCTGCTCTGGGTGGT	3000
Db	2941	TTATCTGCTGTTATTATTCATAAGCTCGGTAGCGAAAAATTCAGTTCTGCTCTGGGTGGT	3000
Qy	3001	ACTGGCGGTATATTACATAGGGATCGGTGTTGCTGGCAGTTATATCCCAAGTTTACGA	3060
Db	3001	ACTGGCGGTATATTACATAGGGATCGGTGTTGCTGGCAGTTATATCCCAAGTTTACGA	3060
Qy	3061	TACCGATTATGGGTGAATGGGAACCTTATGGCTGGCACTGGCGTTCTGCTTGGTGGCG	3120
Db	3061	TACCGATTATGGGTGAATGGGAACCTTATGGCTGGCACTGGCGTTCTGCTTGGTGGCG	3120
Qy	3121	GTGTCATTGGCATGATCTCCTTCGCTCATGTTTAAACGCCCTGGACATATGCAATAATTTAA	3180
Db	3121	GTGTCATTGGCATGATCTCCTTCGCTCATGTTTAAACGCCCTGGACATATGCAATAATTTAA	3180
Qy	3181	CTCCCGCTGAGAAGTTTGCAAGAATTAACTCGGCGAGTAACCTTACTTTATACCAACCGCA	3240

Db 61 CCGTGGCGACATTGTGATCCTGCCAACACGCGTGGATCTCTGACCCCTGGCCAGCCCTGTG 120
Qy 846 TAAGGGCCCAATCATGACAAATAACAAACCGTTATTGGTGTGATGGGATCAGGCAG 905
Db 121 AGAGGATGTGATACGATACGAAACATTAATTTGGCGTGGATGTCTGGGTTCGGCGAG 180
Qy 906 TGTCCGGCGGGATTTTGTGATCAAGGGATCTCTGCTATCCCATGCGCACAGAAAAAT 965
Db 181 CGTGGCGCGGGGTATTCAATCTCTGGGGGCGAGCTCTCCGCCACGCGACCCGGGAAT 240
Qy 966 CACGACTACGGCGCGCAGGAAAGCCGCTGGAACAGTCGACGAGCAGGATCTGGCAGGC 1025
Db 241 AACCCCTGTTTCGAGCGCGCGGAAACAAAGTCGAGCAGTCGACCGGTGAGATCTGGCAGGC 300
Qy 1026 GGTCTGTTCATGTATTCGTAATGGGCTCACCTCTGGCAGACGTTGTGACAAAGTGTGGC 1085
Db 301 GGTGTGCTACTGATTAATAACGCGGTGGCCAAACGCGCGCTTTCGCGCTGCTCGATCGC 360
Qy 1086 AGGCATCGGTTTGTATGCCACTGTCTCTGTGGTACTGGATTAATAACGGTGTATCCAT 1145
Db 361 CGGATCGGCTTCGATGCCACTGCTCGCTGGTGTGATTTGGCGATACGATGGCGCGCT 420
Qy 1146 GCCTGTACGCCGGAAGAGATGCAAGCAAAATATCATTTCTGTGGATGGATCAACCGGC 1205
Db 421 GCGGTGGGCTTCGGGACGATGCGGACCGCAACATTAATCTGTGGATGGATCAFCGCGC 480
Qy 1206 CACGAAACAGCAGAGCGAATCAATGCCACTCACCATCCGGTGTGTAACACTACGTCGGTGG 1265
Db 481 CACGGGCGAGCGGAAAGATCAATGCCACGGGCCACCCGCTACTGCGGTACGTTGGCGG 540
Qy 1266 TAAATTTCCGCTGAAATGGAAACACCGAAATTTCTCTGGGTGAAAGAAATATGCCAGA 1325
Db 541 CAAATCTCGCGGAGATGCAAGCGCGGAAATTTGTGGCTGAAGGAGAAATCGGCCCA 600
Qy 1326 GATCTACGAAGCTGCGGACAAATTTTCGATCTGCGCGATTTCTGACCTGGCGGCTAC 1385
Db 601 CATCTATCAGTGGCGGCCACTTCTTCGACCTCGCGGATTAACCTCACCTGGCGCTCAAC 660
Qy 1386 CGGTGATTAGCGGTTTCAGTATGCACCTGTACCTGTAAATGGAGCTGGCTGGCACATGA 1445
Db 661 CGGCGATGAAGCCGCTCGGTCTGCACGGTGAACCTGTAAATGGACCTACCTGGCGCATGA 720
Qy 1446 AAATCGCTGGGATCAGATTAATTTCCGACCATTTGGCCCTTCGACAGTTAGCGGATGAAGA 1505
Db 721 ACAAGCTGGGATCGCGGTTACTTCCGCGAGATTGGCTCGAAAGAGCTGGTGGATGAAGA 780
Qy 1506 TTTTATTCGCATGTGTCATCATATTGTTTCTCCGGAACACCTTGTGGAATGTTTAAAC 1565
Db 781 TTTTGTCCGCATCGGCCAGCGGATTTGATCCAGCACCCCTCGCGCGAGGGCCTTTG 840
Qy 1566 AGCAAGCCGCGCAGAGATGGGATTAATCCCGGCGCACACTGTCTGCTGTAGGCTTAAT 1625
Db 841 CGCCAGCGCGCAGAGGATGGGCTTCGCGATCGGTACGCGGCTGGCGGTGGGATCAT 900
Qy 1626 TGATGCTCACGCTGGTGGCATCGGTACGGTTGGCGTAGAAGTGGAGCGCTGAACAATCT 1685
Db 901 TGAGGCCACACGCGCGGCATCGGTACCGTGGCGGTGTTAACGGCGCGGTCAACAATAT 960
Qy 1686 CGCGTATGTTTCGGCACTTCTTCATGCACCATGGCATCTACCACTTCTCCCTGTTTGT 1745
Db 961 GGCCTATGTCTTCGGCACTTCTTCCTCGACCATGACCAACCACTCAGGAAGCGGTATTCTG 1020
Qy 1746 ACCGGTGTCTGGGGCGGATTAACAGTGGCATGTTCCAGGCTCTGTGTTAGTTGAAGG 1805
Db 1021 CCCGGCGCTCTGGGGCCCTTACTATTCGGCGATGGTCCGGGCTACTGGCTAAGTGAAGG 1080
Qy 1806 CGGCAAAAGTCTCGGGGAGCAGTATTGACAGCTACTTGATTTCCATCCCGGCTGTTTGA 1865
Db 1081 CGGCAAAAGCCCGCGCGCGCTATCGACCACTGCTGAGCTTCCATCCGCGCGCGCGC 1140
Qy 1866 AGAAGCTCGGAAATGGCACAACGTTGTAATACGCCCTCCCGCTCTGGCTGTGCTGATCG 1925
Db 1141 GGAAGCGCGGAGCAGCGGAAAGCGGAGGCGTCCCTCTACCGGCTCTGGCTGGCGGATCG 1200

Qy 1926 AATCCTGAAAAAAGCGCGCAACCATCAGATGCTGTGCCCCTGGCGAAAGGCTACAGCT 1985
Db 1201 CGTCTGACGCAAGTTGCTTCCCTTCTGAGCGGTACGCTGGCGCGGCTCTGCAGCT 1260
Qy 1986 GGTGCGGAAATTTCTCGGAAATCGGCCCCCTTCGACAGATCTCATGCGAGAGCGGTAAAT 2045
Db 1261 CGTTCGGAGTTTCTCGGTAATCGCGCGCGCTTCCCGATCCCGACGGAAGCGCTGAT 1320
Qy 2046 TTGTGGCTGGGTATGGAGCGAGATCTGATAAATTTACTCGCTGTGATATCGCTGGATT 2105
Db 1321 TGCCGGGTGGGATGGAGCGGACCTGGATAACCTGACCGCGCTGTACGTTGCCGGGCT 1380
Qy 2106 ATGCGGAATTTGTTATGCTCTCGGCAAAATTTCTCACGCTCAAAACAGGCGCAGGAGTAGT 2165
Db 1381 ATGCGGATTTGCTATGCTCTGCGACAGATCATTTGACGCCCGCGCTGCGGAATTGA 1440
Qy 2166 GAGTAAAAATATCGTTTATAGGGGGGTGGCGGAGCATCCACTGGTACGACAATTTCT 2225
Db 1441 GAGTAAAAATATCGTTCATCAGCGCGGAGCGGCGAGCATCCGCTGGTGGCTGAGCTGTT 1500
Qy 2226 GCGAGTACCTCGGTTATTCGGTCAATACCACGCAATGCTCGCAACCGGTTTTATTAGG 2285
Db 1501 GCGCGACCCCTGGCGGCTTACGCTGCTGAGCACCCGCGAGCGCGGCTGCTGTAGG 1560
Qy 2286 CTCGSCCATTTCTTGGTGTCTGCGCGGAAATATTTGCACCTTCTGTTGGCGAAAGCATGCA 2345
Db 1561 TTCCGCCCATCTCGCGCGGTGGCGGCGAGTCTCGGCTTCTGCTGCCGAGGCCATGAA 1620
Qy 2346 ACAATTCACCATGTGGGATAAATATTATTTATCCGCAAGACGCTATCAATCTCTTCATCA 2405
Db 1621 GCAGTTTACTCAGTGGATGCCACCTATCACAGCGAGACGGCTTTAGCCCCCTTCATCA 1680
Qy 2406 TCGTGCATATGAGGCTTATAAGCAGTTGACGATGCTCAAAATTTACTCAGAGACTAAT 2465
Db 1681 GCGCAGGTATGGGCTTATAAGCGCTACACAGCGCGGAGGCTGATTCGCGAATAACA 1740
Qy 2466 AACACGCGCTGAGCGCTGTTTTCAGGCAATCACTAATAACGACTCACTCCCGTAAATATC 2525
Db 1741 TACCTCCCTT-----ATTTTATAACAATAATGCTCCGCTGTGTGCA 1781
Qy 2526 CCGGAGTGCATTCATCTGCACCCCTAAAAACGAGGTTTATATGCTCCAGAAATAATAACAG 2585
Db 1782 CGGGAG-ACACTCACTCTGCACTCTGAAATCGAGGTCACATGTCGGTTAAATAACAAACAG 1840
Qy 2586 TGGTTGGGTTTGGCCACTGCATCTGATATGGGATATATCGCCATCGCAGTATTTATGACT 2645
Db 1841 TGGTACGCTTACCCTCAACCTCATCTGGGATATGTCGCATCGCCGCTGTTATGAGG 1900
Qy 2646 GGTGATGTTTTCGAACCTCGCATTTTATCGCACTATATTAATTCGTTAGGCTTCACACT 2705
Db 1901 GCGCAGCGCTTTGAGCTGGCTTTTCTGTGCGCACTACATCAAAAGCACTCGGATTCACCCG 1960
Qy 2706 GCGGAGGCTCTTTTGGCTTTTACGCTCTACGCGCTGGGGGCTGGCCCTTTCCGCCCTGGGTT 2765
Db 1961 GCGCAGGCTTCTTTGGCTTTTACCCCTCTACGCTTGGCTGGCGCGCTGTCGCGCTGGGTC 2020
Qy 2766 TCTGGGTTAGTAGCGGAAATCATCACGCGCAAAACCATGCTGATTTGGTTTTGTCCTA 2825
Db 2021 TCCGGCGTGGTGGCGGAAATCATTTACCCACGTAAGGCACTGCTGATTTGGCTTCTGCTA 2080
Qy 2826 TGGTGGCTATTCATGTTCTGTTTCTGGTCTGTTGATTAGGACAGGCAACATATGGGTTA 2885
Db 2081 TGGTGTGATTTCCACGCTACTGTTCTCTGCTCTCGGCTCGGTCGCGCCACTATCGCGTG 2140
Qy 2886 ATCCTGCTGTTTACGGGATTCGTTAGCGTATCCGCTATTTCTCTACTCATTTATC 2945
Db 2141 ATCCTGCTGTTTATGGCATCCGCGGCTGGGCTACCCGCTCTCTCTCTCTCTCTCTCT 2200
Qy 2946 GTCGTTATTTATCAACGCTGGGTAGCGGAAATTTCCAGTTCTGCTCTGGGTTGGTACTGG 3005
Db 2201 GTCGCCATTTATCCACACGCTCGCGACGACAGTTCCAGCTCCGCGCTGGGCTGTTCTGG 2260

QY 372 TTTCATGCCAATGCGAGGCGCTTATATTGGCGGCCAGTGGCTGAAGGTGATCCAGATGTC 431
Db 3473 TTCCACGCCAATGCCGCTCCCTATATTGGCGGCCCTCTGGAGGCGCATCCGATGCA 3414
QY 432 TGGGATCGTGTGTTAAATCTGAATATAAATGCGGCTTTTCGCTGTCTCCGTCAGTCCTG 491
Db 3413 TGGGATCGCATGTTGAACCTCAATATAATGCGGCTTTCCGTACCGCTCCATCCGCTTCG 3354
QY 492 CGCATATGATTGCGCAGAGTGGCGGATATAAATTTTACCAGTTCCATCGCGGCGTC 551
Db 3353 CGCATATGATGGAAGCAAGACTGCGACATCATCATGACCACTCTCTGTTGCGGCTC 3294
QY 552 GTTCCGTTATCTGGAACCGCATCTACACCGGCTCAAAATTTTCCGCTTCAGGATTCGTA 611
Db 3293 GTCCCGGTGATATGGAACCGATTTATACGGGTGCAAAATCCCGTGCAGGCTTTTCG 3234
QY 612 CACATACCGCGCGCAGGTTTCTCAATATGCGTSCGTGGTGCAGCA 671
Db 3233 CACACGGTTGCGCGTCAGGTGCGCGCATGGTATTTCGGCTTGGCGGTCTTTCGCGGCG 3174
QY 672 CCAGTAGTCACTGCCCTGCTTGATGACTGGCCAAAGCCAAATGGAAGGCCCTTGCA 731
Db 3173 CCGGTCTCACTGCGCTCATCAGCAGACTGCGCGCAAGCCAAAGCTCGATGAGCGCTGGCG 3114
QY 732 AATGTTAGCTGATGCAACCGATTTGAAGTGGCGGAATCAGTATTGTTTATGTTGACCCGC 791
Db 3113 GCAGGAGCGCTGATGAGCGGAAAGAGTTGCGGAATCCGTGATGTTTCACTGTCACACG 3054
QY 792 TCGAAAATGTCAACGTCGAGATTTAGTATCTCCCTGGCAGTGTGATCTGTAAGG 851
Db 3053 CCGCGCAATATCACCATCCGCATCTGGTATTCTTCGCGAAGTCTCGATCTCTAAATG 2994
QY 852 CGCAATCATGACAATAACCAAAACCGTTATTGGTTGATGTTGGATCAGGCAAGTCCG 911
Db 2993 GG----AGCGACCATGACCAACTATTATCTGGCGCTGATGTCGCGACCGAAGCGCG 2938
QY 912 CGCGCGGATTTTGATCTCAACGGATCTCTGATATCCCATGCCACAGAAAATACAGAC 971
Db 2937 GCGCGGCTGTTCGACGCGGTGGCAACATGCTTGATCGGCGAGCGGATATGCCCAT 2878
QY 972 TAGCGCGCAGCGAAGCGGTGGAACAGTCCAGCAGGAGATCTGGCAGCGGCTG 1031
Db 2877 ATGCGCGAGCGCGGCTATTGTCAGCAGTCCAGCAGCATATATGCGAGCGCTTTG 2818
QY 1032 TTCATGTTATTCGATGCGCTCACTCTGCGACAGCTTTGTCACAAAGTGTGCGCAGCAT 1091
Db 2817 TGAAGTGTTCGCAAGTGTGAGGTTGCTGGCGCTGATCCGACGCGCTTTCGCGCAT 2758
QY 1092 CGSTTTTGATGCCACCTGTTCTGCTGCTGATGGAATAAAACGATGATCCATGCTCT 1151
Db 2757 CGGCTATGATGCCACCTGTTCTATTTGCTGCTGGCGAGGCGCAACCGCTTGGCT 2698
QY 1152 CAGCCCGAAGGAGATGCAAGCAAAATATCATTTGTTGATGATGATCAGCGGCCACCGA 1211
Db 2697 TGTCTCTTCAAGATGCGCGCGCAACATCATCTGTTGATGATCATCGCGAGTGA 2638
QY 1212 ACAGCAGACGCAATCAATGCGCACTCACCATCCGGTGTGAATAGTCTGGTGGTAAAT 1271
Db 2637 GCAGGACAGCGCTATCAATACCAACAGCGGATGTGCTGGGTTATGTCGGCGGCGCAT 2578
QY 1272 TTGCGCTGAATGGAACACCGAAATCTCTGGCTGGAAGAAATATGCCAGATCTA 1331
Db 2577 TTGCGCGAAATGGAACCTCCGAAATCTCTGGCTGGAAGAACACAGCGCGAAACGTT 2518
QY 1332 CGAACGTGCGGACAAATTTTCGATCTGGCGGATTTCTGACCTGGCGGCTACCGTGA 1391
Db 2517 TGCTGCCGCTGGCAGTTTTCGATCTGACCGATTTTCTACATGGAATCTCTGGGCA 2458
QY 1392 TTATAGCGCTTCAGTATGCACTGTTTACCTGTAAATGGAGCTGCTGGCAGATGAAATCG 1451
Db 2457 TCTTGCGCGCTGGCGCTGTCAGCTAACCTGCAATGGACCTATCTCAGCCATGAAAGCG 2398
QY 1452 CTGGGATCCAGATTTATTTCCGCAACCATTTGGCCTTCAGAGTTAGCGGATGAAGATTTAT 1511

Db 2397 TTGGACGAGACCTATTTCGCGCAGTTCGCGCAACTCGCGGATGAAAGTTTCGT 2338
QY 1512 TCGCATGTGTCATATTGTTTCCCGGAACACACCTTGTGGAATGTTTAAACAGACA 1571
Db 2337 CGCATCGTACAGAGTTTCGCGCGGTGCGGCAAAATCTTGG---CGGTTGAGCAACA 2281
QY 1572 AGCCGCGCAGAGATGGGATTACTCCCCGCGCACACCTGCTGTAGGTTAAATGATGC 1631
Db 2280 GCGCGCAGCGAATCGGCTTGAGCGCGGTACGCAATTCGCGCGGTCTGATCATGC 2221
QY 1632 TCACGCTGTGGCATCGGTACGTTGGCG-----TAGAAGTGGAGCGGTGAACAA 1682
Db 2220 ACATCGCGCGGTATCGGCACGTTGGCGCGCGGTTCGGAAGTTCGATATTTCGCG 2161
QY 1683 TCTCGGTATGTTTTCGCGCACTTCTTCATGCACATGGCATCTACCACTTCTCCCTCGTT 1742
Db 2160 TATGGCTATGTTTCGCGCACATCGGCTTGACCATGACCACTGACAGCGCTGTTT 2101
QY 1743 TGTACCGGCTGTCTGGGCGCTTATACAGTGGATGGTTTCCAGTCTGTGGTTAGTTGA 1802
Db 2100 CGTGGATGCTGTGGGCGCTTATTTCTCGGCACTGGTTCGCGGCTGTGGTTGAACGA 2041
QY 1803 AGCGCGGCAAGTGTCTGGGAGCAGTATTGACAGCTACTTGTATTTCCATCCGCGCT 1862
Db 2040 AGCGCGCAATCGGCTGCGGTTGCGCTATTGATCATCTGATTCACATGCATCCCTTCGC 1981
QY 1863 TGAAGAAGCTCGCAAAATGGCAACAGTGTGAATCAGCCCTCCCGCTTGGCTTCTGA 1922
Db 1980 GCGGGAAGCGGAAGGAGCGCGGATCAGGCAAGGGCTGGCAGACAGCTCCTCGTC 1921
QY 1923 TCGAATCTCGAAAAACCGCGCAACCATCAGATGCTGTGCCCTGGGCAAGGGGTACA 1982
Db 1920 CGAAG---TTGAAGCTAGAGTGTCTGAAAAAAGCGCATGATCGTTGGCGACATTC 1864
QY 1983 CTGTGTGCGGAAATTTCTCGGAATCGCGCCCTTCGAGATCCCTCATGCCAGCGGT 2042
Db 1863 TGTGTGCGCGAATTCCTCGTAACTGTCACCTTCGCGGATCTGTGCGGCTGCGGT 1804
QY 2043 AATTTGTGCGCTGGGTATGAGCAGATCTGGATAATTTACTGCGCTTGTATATCGCTGG 2102
Db 1803 TATTGCGGCGTGGACCTTGATACAGGATGGACAGTCTGGCGGCACTTTATCTCGCGG 1744
QY 2103 ATTATGCGAATTTGTTATGCTGCGCAAAATCTCAGAGCTCAAAACAGCGGAGGT 2162
Db 1743 GCTTTGCGGCTGGGTATGTTGTCGCGAGATTTATCGAAGCGCAGCGGCGCAAGGTAT 1684
QY 2163 AGTGAATAAAATATCGTTATTAGCGCGGTGCGCGGAGCATCCACTGTCAGCAAAAT 2222
Db 1683 TGTGACTGACATATTGTTGTTGCGGCGCGCGGCTCAAAATCTCGTGGCGGAGGT 1624
QY 2223 TCTGCGCATACCTGCGGTATTTCGCGTATTACACGCAATGCTGCAACCGGTTTATTT 2282
Db 1623 GCTCGCGATGCGACGCGCTCGTTGTTACCGCTTCGACATCGCGGAGCGGTTCTGCT 1564
QY 2283 AGGCTCGGCGCATTTGTTGCTGTCGCGGAAATATTGACACTTCTGTTGGCGAAGCAT 2342
Db 1563 CGGCTCGGCGATTTAGTGTGCGGTTGCGTGGTATTCCCGATCTTGTGACGCGTAT 1504
QY 2343 GCAACAATTCACCATGTCGATAAATATTATTATCCGGAAGAGCTATCAATCTCTTCA 2402
Db 1503 GCAAGTCATGTCGGAATTTAGGCGCAGCAACAGCGCGGAGCGCGGTCGCAAAATGGCA 1444
QY 2403 TCATCGTCATATAGGCGCTATAAGCAGTTGACGATATCTGC 2444
Db 1443 TGACCAACCGTTTGGCGCTTCATGTTGCTTCAGGCGACTGC 1402

RESULT 5

AF045245/c

LOCUS

DEFINITION

5930 bp DNA linear BCT 30-AUG-2001
Klebsiella pneumoniae D-arabinitol transporter (dalT),
D-xylulose kinase (dalK), D-arabinitol dehydrogenase (dalD), and

QY	3081	GGAACTTATGGCTGCACATCGGCTTCTGCTTTGGTGGCGGTGTCATGCCATGATCTCC	3140
Db	1182	GGCACCCTGTGGCTGGCGCTTCTGCTTACCGGGGGGTGATGCCCTGGTTTCC	1123
QY	3141	TTGCGGTCATGTTAAACGCGCTGGACATATGCATAATTTAACTCCCGCTGAGAAAGTTTGA	3200
Db	1122	CTGCGGCATATTCAAACGCGCGAGCATATGCAGAACTCTCACCACCGCTGAGAAAGTTTCC	1063
QY	3201	GAATTAAAGTCGGGACGACTTACTTTATACCAACCGCAATATTTTCTCTCCAGTATT	3260
Db	1062	GAGCTGGAGCGCGCGAGCTGCTCTATACCAACCGCAATATTTCTGCTGCGACATG	1003
QY	3261	GTGCGCATTTAATAACCTTATCGTTATTCGGTFTTTCGGGTCAATATGCCAATGATGTTT	3320
Db	1002	GTGAGATCATCAATACCTCTCGCTTGTTCGGCTTTTTCGGGTGATATGCCGATGATGTTT	943
QY	3321	GTGGATGAACCTGGGATTCACCACTCTGAATGTTTCAGCTGTGGCGCATTTTCTTT	3380
Db	942	GTGATGAGCTGGCTTTCAGCACTTCAGAAATGCTCCAGGTATGGGCACTCTCTTCTTCTTC	883
QY	3381	ACCACATTTTCTCTAATATTTTGGGGGATTTGGGCAGAAATAATGGCTCGATCGCT	3440
Db	882	ACCACCATTTTCTCTAAGCTGTGTGGGGATCCTCGGGGAAAACTGGCTGGATGAAG	823
QY	3441	GTTATTCGCTGGTTGGTTCCTCGGGATGGCAGCATCAAGTTTACGCTTTTACTACATG	3500
Db	822	GTGGTGGCTGGTTCGGCTGCACTGGGATGGCGCTCTCCAGCTTGGCGTTTACTACATA	763
QY	3501	CCGCAATCTTTGGTCACAACTACTGCAATGCAATCTTCGGCGATTCCTCTGGGAAC	3560
Db	762	CCTCAGCATTTGGCATAGCTTTGGCATGGCGCTATCCCGCATCGCCCTCGGGATC	703
QY	3561	TTTGTGCTGCATTTGTCGCGATGGCGCTGTCTTCCCGGCATCGGAACAAACAAA	3620
Db	702	TTTGTGCGCGCTTCTGTCGCGCTGGCGCGCTGTCTTCCCGCGCTGAGCGAAGCATAA	643
QY	3621	GGTCTGCAATCTCGTTTACAACTCTCTCGGGTATGTCCTAACCTTCTCGCTCGCGA	3680
Db	642	GTGCGGCCATTTTCGGCTCTATAACCTGTTCGGGGGATGTGCAACTTCTCGCGCGG	583
QY	3681	ATTGCGCGTGGTATTATACCGTGGTTTACGATATCGGTGTGTCATTCGCTATACGCA	3740
Db	582	ATTGCGCGTGGTGTGCTGCGCTTCTTCAGACCATTCGCGGTGTGATCGCCCTATACGCA	523
QY	3741	TTGATCTATTATGGCCCTTTCTCTTTCGCGCATTCATTCGGGTTTGAGCAGCAGGATTC	3797
Db	522	TTGTACGTTGGTGGCCCTTTTCTCTCTCGCGCTTATTCGCGGTGGAGCAGCGGGTTC	466
RESULT 6			
AE008252/c			
LOCUS	AE008252	10295 bp	DNA linear BCT 18-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 56 of 187 of the complete sequence.		
ACCESSION	AE008252 AE007870		
VERSION	AE008252.1 GI:15158917		
KEYWORDS	Agrobacterium tumefaciens str. C58 (Cereon).		
SOURCE	Agrobacterium tumefaciens str. C58 (Cereon)		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
REFERENCE	1 (bases 1 to 10295)		
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.		
TITLE	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 10295)		
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA		
COMMENT	Approximately 800 bp of telomeric sequence missing from the left		

FEATURES		end of the chromosome and 200 bp missing from the right end.
source		Location/Qualifiers
		1..10295
		/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
		/strain="C58"
		/db_xref="taxon:181661"
		/complement(195..704)
		/gene="AGR_L_1070"
		complement(195..704)
		/gene="AGR_L_1070"
		/note="hypothetical protein"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1070p"
		/protein_id="AAK89112.1"
		/db_xref="GI:15158918"
		/translation="MAAEKRGVPHINERPAHGRALMKSGEMSVTMLGHLFOLIL
		PRGRHVDVAFIAHEVKAERDRLRADAETADIDGGGGSDTMDMWNADLMV
		VCAINGCPFEHRRGGEFRNEAYVIGMVHCSSCLGLHRTNSTRKAPFHPILQSV
		EVGYSFSSVHCWGS"
		747..2297
		/gene="AGR_L_1071"
		747..2297
		/gene="AGR_L_1071"
		/note="PROLINE/BETAINE TRANSPORTER (PROLINE PORTER II) (PPII)"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1071p"
		/protein_id="AAK89113.1"
		/db_xref="GI:15158919"
		/translation="MVRHRSQQNLNQIQASPLSIDVTVIDDKVKAVTAAALGN
		AFQKDFGVGVFVAVGVKFPFPAAPVQVAAALATFSPFLIRPLGGIFTCAMGDK
		EMRQVLSLTIIIMAASTFCIGLPGYATIGWAPILLLCKLVGVSFGGYTGAAI
		FVAEYAPDRKRYLGSWLDGSIAGFVLGAGLVLLSSAMGEAKFLDWKWRIPFLAA
		PLGLGLYLRUAETPAFTLRQAEEDRQSLKERPMVPIAEIAREHFRSLAICIG
		MYLVNTVITYMLLTPTLTKLNYSEDHVGLIIIAVMWMLFVQPAIGLSDRIGR
		KPFLAAGSTAILFLSLPAFLIASGVQVIFGGLLILAVALNCLIGINASTLPALFPA
		RIRYGALAFNISIIVAGLTPTVTVALVEVTENTYMPAYLMAVAVGIVTTFLEKE
		TANRPLVGDPTNASSRKKAESLLAEQYQHIKESVBEIDELAKLEQKQLQTKREG
		LVGRHDLT"
		complement(2383..3195)
		/gene="AGR_L_1073"
		complement(2383..3195)
		/gene="AGR_L_1073"
		/note="carboxyl reductase"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1073p"
		/protein_id="AAK89114.1"
		/db_xref="GI:15158920"
		/translation="MRRVPEFAPAIIDRQSLGRYLWPNONNDSMOPSGKSVITGAGK
		GIGRACAOILMAARGAEVVAISRTQSDLSLKEIGRSIRVDLADIAATRAAMAEAGP
		CDFLINAGINVLVSVDMSDAGYEAIVGINLRALVTCQEFARGRVANGGCAIVNI
		TSIAGHRGFQDHLCYAASKAGLEGATRLAKELGPHGIRVNAVAPITITULEAEAWN
		DPEKSPQMVRHPLNRFAEAEVAQSIALLLSDSKMISGAVLPVDGGFLAV"
		complement(3158..4768)
		/gene="AGR_L_1075"
		complement(3158..4768)
		/gene="AGR_L_1075"
		/note="D-ribulokinase (EC 2.7.1.47) - Klebsiella pneumoniae"
		/codon_start=1
		/transl_table=11
		/product="AGR_L_1075p"
		/protein_id="AAK89115.1"
		/db_xref="GI:15158921"
		/translation="MASVLVGVDTGTSARAGVFDVAGKLLATARKRPISMHRDGGIA
		EQSSAEVWQAVCDVSVDVSRAGIDPAEVTGIGDFATCSLVVRGPDVTLPVGAADHP
		ERDITVMDHRAVEQAEIRINAGKAVKYVGRISPEMOTPKLLLSNRPDIYARAE
		HFEDLTDLTWKASGALDRSACTVCTKNTYLAHENRWDAEYFTKIGLDLARQGFRII
		GESVYHPCTALGNGLTEGAANKGLVAGTAVAGLIIDAHAGGVGTVAAGGDSRCLGY
		VFTSSCTMTTITTEPAFVFGVWGPYISAWPGAWLNEGQSAAAGAIIDLVLQHPAFA

Db 4878 ACCAATCTCGACCGCGTTTAAATAATCTCTGGAGCGCGGAAGAACCGGAAACCGGTTCCCG 4819
QY 878 ----- 877
Db 4818 CGCTTCAACCTGTTCAAGCACCGGATGATCCGCTCCGAGGAGTATCGCAAAATGGCTTCCT 4759
QY 878 ---TTATTGGTGTGATGTGGGATCAGGCAGTGTCCGCGCGGGGATTTTGTGATCTCAACG 934
Db 4758 ATCTTGTGGCGTCGATGTGGGACCGCTCGGCGCGCGCGGTGTCTTTGTGATGTCCGG 4699
QY 935 GATCTCTGCTATCCCATGCCACAGAAAAATACGACTACGCGCGCCAGCGGAACCGG 994
Db 4698 GCAAGCTTCTCGCCACCGCCAGCGCCGATCAGCATGACACGGGAGGATGGCGGCTTG 4639
QY 995 TGGACAGTCCAGCCAGGAGATGTGCAGCGGCTGTCTCATGTATTCGTAAATGCGCTCA 1054
Db 4638 CCGAGCAATCGAGCGCAGAGGTGTGCGAGCGGTTGTGATAGCGTTCCGACAGCGTTT 4579
QY 1055 CTCTGGCAGCGTTTGTGCACAAAGTGTGGCAGGCATCGGTTTTGTGATGCAACCTGTTCTC 1114
Db 4578 CACGCGCGGATTCGATCGCGCGGAGGTTACCGGCAATTGGCTTCGACGCGACCTGTTCTC 4519
QY 1115 TGGTGTACTGGATAAAACGGTGATCCATTGCCTCTCAGCCGGAAGGAGATGCAAGC 1174
Db 4518 TTGTGGTTCGCGACCCGGTGACGTGACCTCGCGGTAGTGTGCTCGCATCATCCGGAAC 4459
QY 1175 AAAATATCATTTGTGTGATGGATCACCGCGCCACCAAGCAGAGCAATCAATGCCA 1234
Db 4458 GTGATATCATCGTCTGATGGACCAACCGTGTCTGACGAGGCTGAACGCATCAACGCGC 4399
QY 1235 CTCACCATCGGTGTGACTAGTCTCGGTGGTAAATTTGCGCTGAAATGGGAACACCGA 1294
Db 4398 GCAAGCATCGCTCTGAAATATGTGGCGCGCATTTTCAACGGAAATGCAGACGCCA 4339
QY 1295 AATTTCTCTGGCTGAAGAAAAATATCCAGAGATCTACNAAGTGGCGGACAAATTTTCG 1354
Db 4338 AGCTGTGTGGCTTCAGGAAAAAGCGCGGATATCTATGCCGCTGCCGAACACTTCTCG 4279
QY 1355 ATCTGCGGATTTCTGACCTGGCGGCTACCGGTGATTTAGCGGCTTCAGTATGCACATG 1414
Db 4278 ACCTGACGGATTTCTGACCTGGAAGGCTCCGGGCGCTTGATCGCTCCGCGCTGCACGG 4219
QY 1415 TTACCTGTAAATGGACGTGGTGGACATGAAATCGCTGGGATCCAGATTAATTTCCGCA 1474
Db 4218 TCACCTGCAATGGACATATCTCGCCACGAAACCGCTGGGATGGGAATATTTCCACCA 4159
QY 1475 CCATTGGCCTTCGAGAGTTAGCGGATGAAGATTTTATTCGATTTGTCATCATATTTT 1534
Db 4158 AGATCGGCTTGGCGATCTGGCGGACACGGCTTTCGCGCATCGGTGAAGCGTGTCC 4099
QY 1535 CTCGCCGAACACTTGTGAAATGGTTTAAACAGCAACCGCGGAGAGATGGGATTAC 1594
Db 4098 ATCCCGCACCGCACTGGGTGAATGGCTGACCCAGAGCGCGCAAGGCGCATGGGTCTGG 4039
QY 1595 TCCCGGCACACCTGTCGCTAGGTTAATGTATGCTCACGCTGTCGTCATCGGTACCG 1654
Db 4038 TCGCAGGAACCGCGTGGCGCTGGGCTGATCGACCCCATCGCGCGGTGTGCTGCTACCG 3979
QY 1655 TTGGCTGAGAAGTGGAGCGCTGAACAAATCTCCGCTATGTTTTCGGCACTTCTTCATGCA 1714
Db 3978 TGGCTCGGGTGGTATGATCATCAGATGCTCGGTTATGTTTCGGCACCTCGTCTCGCA 3919
QY 1715 CCATGCGATCTACCACTTCTCCCTGTTTGTACCGGGTGTCTGGGGCGGATATTACAGTG 1774
Db 3918 CCATGACGACCACTGACCGCGCTTGTGCGGGTGTCTGGGGTCCCTATTATTCG 3859
QY 1775 CGATGTTCCAGTCTGTGTAGTTGAAGCGGGCAAGTGTGCGGGAGGACGATTTTG 1834
Db 3858 CAATGGTTCTGGCGATGGCTCAATGAAGGGCGGACGTCCTCGCGGTGCGGCGCATGG 3799
QY 1835 ACCAGCTACTTGATTTCCATCGCGGTGTTGAAGAACTCGCGAAATGGCACACGTTGTA 1894
Db 3798 ATTACCTTGTGAGTTGACCGCGCTTTGCCGAGCAAAAGGCTTCCGACAGGACG 3739

QY 1895 ATAGCCCCCTCCCGCTGCTGCTTGTGATCGAATCTCTGAAAAACGGCGCAACCATCAG 1954
Db 3738 GCAAGCATTTGCCAGTCTGGCTTGCAGCGCGCTCAGCCTCGCGCATCGCTTCCG 3679
QY 1955 ATGCTGTCCCGCTGGGAAAGGCTACAGTGTGCGGAAATTTCTCGGAAATCGGCGCC 2014
Db 3678 CGCAGTGAAGCTTGGCGAAGATTTTATATGTCGCGAAATTTCTTGGCAATCGGCGC 3619
QY 2015 CTTCCAGATCTCATGACAGAGCGGTAATTTGTGGCTTGGGTATGGAGCAGAGATCTGG 2074
Db 3618 CTTTCCCGATCCCATGCTCGGCCCATATTGCCCGGTTATGGCATGGAGACGGGCTGG 3559
QY 2075 ATAATTTACTCGCTTGTATATCGTGGATTTATGCGGAATTTGGTTATGTTCTCGCCAAA 2134
Db 3558 ATTCCCTCTCGCGCTCTATGTGGCGGGCTTCTCGCCTCGGTTACGGGCTTCGCCAGA 3499
QY 2135 TTCTCGACGCTCAACAGCGCAGGAGTAGTAGTAAAATATCGTTATTAGCGGCGGTG 2194
Db 3498 TCATGAAAAACAGCGGCCGACGCGCGGCGGAAACCATCAGCGTCAGTGGCGGAG 3439
QY 2195 CGGGCAGCATCCACATGGTACGACAAATTTCTGCAGATACCTCGCGTATTTCCGCTCATTA 2254
Db 3438 CAGGCGCTCATCGCTTGCAGGACGTTCTGCGCATGGACCGGCTCCCGCTTGAAC 3379
QY 2255 CCACGCAATGCTCGCAACCGGTTTTATTAGGCTCGGCCATTTCTTGGTGTGTCGCGGAA 2314
Db 3378 TGACCGAATGCGAGGACCGGCTGCTGTTGGATCTGCCATGCTCGCGCGTTGCGCGCTG 3319
QY 2315 ATATTGACCTTCTGTTGGCGAAGCATGCAACAATTCACCCATGTGGATTAATATTT 2374
Db 3318 GAACCTATTCGGACCTGATGGCTGCCATCGCGCGGATGTCGCTATCGACAGCAGTGCA 3259
QY 2375 ATCCGCAAGACGCTATCAATCTCTTCATCATCGTCGATATAGGCTTAACGACGTGC 2434
Db 3258 TGCCGATCCGATTTTCAGAGATTCACAGCGGCTTACGACGCTTCTTGGCGCTG 3199
QY 2435 AGCATACTGC 2444
Db 3198 AGAATCGCGC 3189

RESULT 7
AE009360
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AE009360 11437 bp DNA linear BCT 20-DEC-2001
Agrobacterium tumefaciens str. C58 linear chromosome, section 130
of 187 of the complete sequence.
AE009360 AE008689
AE009360.1 GI:17742779
Agrobacterium tumefaciens str. C58 (U. Washington).
Agrobacterium tumefaciens str. C58 (U. Washington)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.

1 (bases 1 to 11437)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M.,
McLellan, E., Palmieri, A., Raymond, C., Rouse, G., Paulsen, I.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Biddle, P., Jung, M.,
Krespan, W., Parry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
and Nester, E.W.

The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
MEDLINE
21608550
PUBMED
11743193
2 (bases 1 to 11437)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,

Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M., McLelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V., and Nester, E. W.

TITLE Direct Submission
JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

FEATURES

Location/Qualifiers
1. .11437

/organism="Agrobacterium tumefaciens" str. C58 (U.

Washington) "

```

        /strain="C58"

```

/db_xref="taxon:18083

QY 1775 CGATGGTCCAGGCTGTGGGTAGTGAAGGCGGCAAGTGTCTCGGAGACGACTATTG 1834
 Db 9890 CAATGGTCTCGGCGATCGGCTCAATGAAGGCGGAGTGTCTCGGCGCGGCGGCGCATCG 9949
 QY 1835 ACCAGCTACTTGAATTCATCCGCTGTGAAGAGCTCGGCAATGSCACACAGTGTGA 1894
 Db 9950 ATTACCTGTGCGAGTGTACCGCGGCTTGGCGAGGCAAGAGCGTTCGCGCAAGAGGACG 10009
 QY 1895 ATCAGCCCTCCCGCTGTGGTGTGTGATCGAATCCCGAAAAAGCGCGCAACCATCAG 1954
 Db 10010 GCAAGAGCATGCGAGCTGTGCTGCGAGCGCGGCTCAGCTCGCGCATCGGCTTCG 10059
 QY 1955 ATGCTGTGCGCTCGGCAAGAGCTACAGTGTGCGGGAATTTCTCGGAATCGCGCC 2014
 Db 10070 CGCGAGTGAAGCTTCCGGAAGATTTTCATATGTTGCGGGAATTCCTTGGCAATCGCGCG 10129
 QY 2015 CTTTCGAGATCTCATCCAGAGCGGTAATTTGCGCTGGGTATGAGGAGGAGATCTCG 2074
 Db 10130 CTTTCGAGATCCCATCTCGCGCATCATTTGCGGTTATGGCATGGAGAGGCGGTGG 10189
 QY 2075 ATAATTTACTCGCTTGTATATCGTGTGATATCGGGAATTTGTTATGTTGCGCCAAA 2134
 Db 10150 ATTCCCTCGTGGGCTCTATGTGGGGGCTTCTCGGCTCGGTTACGGGCTTCCCGAGA 10249
 QY 2135 TTTCTGAGCTCAACACAGCGGAGTAGTGAATTAATAATATCGTTATAGCGCGGTG 2194
 Db 10250 TCATCGAACACAGCGCGCAACGCGCGGTTGGAAACCATCAGCTCAGTGGCGAG 10309
 QY 2195 CCGGCGAGCATCCACTGTACGACAAATTCGGCAGATACCTGCGGTATTCGGTCAATTA 2254
 Db 10310 CAGGCGCTCATCCGCTTCCAGGAGTGTGCTGGCAGATGCGACCGGCTCCCGGTTGAAC 10369
 QY 2255 CCACGCAATGTGCGAAGCGGTTTATTAGGCTCGGCGCATTTCTGCTGTGCGCGGAA 2314
 Db 10370 TGACCGAATGCGAGCGGCTGTGCTGTGATCTGCCATGCTCGCGCGGCTGCGCGTG 10429
 QY 2315 ATATTGACCTCTCTTGGGAGAGGATGCAACAAATCACCCATGTTGATTAATATTATT 2374
 Db 10430 GAACCTATCCGAGACTGATGCTGCGATGCGCGGCTGCGGTATCGACAGCAGTGCA 10489
 QY 2375 ATCCGCAAGACGCTATCAATCTTTCATCATCTGATGCGATGAGGCTTATAGCAGTTC 2434
 Db 10490 TGCCCGATCCCGATTTTCAGAGATTCACGAGCGGCTTACGACGCTTTCGCGGCTGC 10549
 QY 2435 AGCATACTGC 2444
 Db 10550 AGAATCGCGC 10559

RESULT 8
 KPNRBTDK
 LOCUS
 DEFINITION K.aerogenes ribitol dehydrogenase (rbdD and rbtK) genes, complete cds and 5' end respectively.
 ACCESSION M25606
 VERSION M25606.1
 KEYWORDS ribitol dehydrogenase.
 SOURCE K.aerogenes DNA, clones lambda-[prbt, prbt, dal] and pJCW1.
 ORGANISM Klebsiella aerogenes
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Klebsiella.
 REFERENCE 1 (bases 1 to 851)
 AUTHORS Loviny, T., Norton, P.M. and Hartley, B.S.
 TITLE Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the structural gene
 JOURNAL Biochem. J. 230 (3), 579-585 (1985)
 MEDLINE 86050424
 PUBMED 2933028
 FEATURES
 Location/Qualifiers
 1. 851
 /organism="Klebsiella aerogenes"
 /db_xref="taxon:28451"
 28. .777
 CDS

/note="ribitol dehydrogenase rbdD"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25140.1"
 /db_xref="GI:149319"
 /translation="MKHSVSMNTSLGKVAATGAASGIGLECAARTLLGAKVLLI
 DREGEKLNKLVAELEGNALFAQVLDQMADQVNDLQGLTQLTGRDIDPHANAGYIG
 PVAEGDPDWDNRVHLNINAAFRVSRVPLHIAQKSGDIIFTSSIAGVVPIWEPVY
 TASFVAQAFVHTTTRVOAQQYGRVAVLPVPPVTTALLDDWPKAKMDEALANGSLMQP
 IEVAESVLEFVHTRSKNVTVRDIVILPNSVDL"
 804. .851
 /note="ribitol dehydrogenase rbtK"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25141.1"
 /db_xref="GI:149320"
 /translation="MHNDQNIIGVDVGS"
 172 a 248 c 264 g 167 t
 BASE COUNT 226 bp upstream of SalI site.
 ORIGIN
 Query Match 11.8%; Score 468.6; DB 1; Length 851;
 Best Local Similarity 75.9%; Pred. No. 1.3e-125;
 Matches 579; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
 QY 86 AAAGGATTATATGATGAATCACTCTGTGCGCTCTATGAATCTCCCTTAATGCGAAAGT 145
 Db 15 AAAGGATTATAAATGAAGCACTCTGTCTCTCTATGAATCTCCCTCAGCGGTAAGT 74
 QY 146 TGCACCTATCACTGCGCTGCGTCAAGTATTGGCCCTGCAATGTGCAAAAACCTGCTCGA 205
 Db 75 CGCCCGCATCAACCGCGCGGCTCGGATCGAGTCCGCGGAGGACCTCGCTTGG 134
 QY 206 TGCAGGAGCAAAAGTAGTATTGATTGACCGGGAAGCGACAACTGCACAAAGATTGTCGC 265
 Db 135 GGCTGGCGCAAAAGTAGTACTGATCGACCGGGAAGGCAAAAGCTCAACAAACTGGTCGC 194
 QY 266 TGAAGTAGGCGAAAACGCGCTACGCGCTGCAACTGCGATCTCTTCAATAATCAGCAAGTCGA 325
 Db 195 CGAACTTGGCGAAAACGCGCTTCCGCTTCAGGTGCGAGTGCAGGCGGACCAAGTCCA 254
 QY 326 TAACATGCTGGCGGACATATCGAACTGGCGGGTGGGATATTTTCATGCCAATCGC 385
 Db 255 TAACCTACTGCAAGGCAATTTTTCAGCTTACCGGGGCTCTCGATATCTTCCAGCGCAACGC 314
 QY 386 AGGCGCTTATATTGGCGGCCAGTGGCTGAAAGTGATCCAGATGCTGGGATCGTGTGT 445
 Db 315 CGGCGCTTATATCGCGGGCGGCTGGCGAGGCGGACCGGAGCTGGGACCGCGTGT 374
 QY 446 AAATCTGAATATAATGCGGCGTTTTCGCTGTGCTGCGAGTCTGCGCGCATATGATTGC 505
 Db 375 GCACCTTAATATCAACGCGCGCTTCCGCTGCGTGGCGAGCTCTGATCGC 434
 QY 506 GCAGAGTGGCGGATATAATTTTACAGTTCCATCGCGGGGCTGTTCCCGGTATCTG 565
 Db 435 GCAAAAATCCGGGGATATATCTTCCAGCTCGATCGCGGGGCTGGTGGCGGTGATCTG 494
 QY 566 GGAACCGATCTACACCGCGTCCAAATTTTCGCTTCAGGCAATCTGACACACTACCGCGC 625
 Db 495 GGAGCTGTCTATACCGGCTCGAAATTCGCGGTGAGGCAATTTGTGCACACCACTGTCG 554
 QY 626 CCAGTCTTCTCAATATGCGGCTGCGGTGCGGTGCTGCGAGGACCACTAGTACCTGC 685
 Db 555 CCAGTGGCGGATATGCGTGGCGCTGGCGCCGCTATTTCGCGGCGCGCGGTGTTACCGC 614
 QY 686 CTTGCTGATGCTGCGCAAAAGCCAAATGGAAGAGCCCTGGCAAAATGGTAGCTGAT 745
 Db 615 TCTGCTGACGACTGGCCAAAAGCCAAATGGACGAAGCGCTGGCAACGGCAGCGCTGAT 674
 QY 746 GCAACCGATTGAAGTGGCGGGAATCATGTTTATGTTGACCGCTCCAAAATGTCAC 805
 Db 675 GCAGCGGATTGAAGTGGCGGAGTGGTGTCTGTTATGTTGAGCGGCTTCAAAAACGTCAC 734
 QY 806 GTCGCGAGATTAGTGTATCTGCTGCGCTGGCAGTGTGCTGTCTGTAA 848

DTALPVKTLSCGNAQRTAIAKWATQPRILLDSPTVGVGDIANKEGYIHTAKALAEQ
 MAVLMICDEIPEAYINSHRVLVMRKGLIAIAEFYPHQCTEQQIAEYVWG"
 /gene="y0235"
 complement(6021..7079)
 /gene="y0235"
 complement(6021..7079)
 /gene="y0235"
 /function="transport; transport of small molecules;
 carbohydrates, organic acids, alcohols"
 /note="residues 25 to 350 of 352 are 53.37 pct identical
 to residues 5 to 327 of 328 from GenPept :
 >gb|AA654669.1|AE005211.4 (AE005211) putative periplasmic
 binding protein, probable substrate ribose [Escherichia
 coli O157:H7 EDL933]"
 /codon_start=1
 /transl_table=11
 /product="putative periplasmic binding protein of ABC
 transporter"
 /protein_id="AAM83829.1"
 /db_xref="GI:21956911"
 /translation="MQHNKISYLAIVPLPERKQEGTIAMKFNALLNNVIVSACMLFTT
 QTLAEKKHETIAVYAKYVTPWFTRMEVGNKAEAKLNINAYQVGATPDPAQQVKVI
 EDLTKAIVDAIIVVPNDKAVLEPVLKKAQEGIVLTHSPDQRIQGDWVETIDSKY
 AOAANDELKAMGNGGVAIVGSLTVPLHNADVAIYKQEKYPMFEVTPRLPVA
 ENIDKSYTFLDLMTKTYPOKGGIIGFSGSLPGAGQAVAKKRAKDQIAVVGIAIAPAQA
 APYLMRGDIKALLWDPNDAGFAVEIANGLNQKRVTELTIDGLGKADVDSKNGVI
 RFNKILEVTYKDNKTLGF"
 /7223..8724
 /gene="y0236"
 /note="y0236"
 /7223..8724
 /gene="y0236"
 /function="putative factor"
 /note="residues 25 to 333 of 333 are 66.66 pct identical
 to residues 20 to 327 of 328 from E. coli K12 : B3378;
 residues 28 to 333 of 333 are 66.66 pct identical to
 residues 1 to 305 of 306 from GenPept : >emb|CAD07638.1|
 (AL627274) putative acyltransferase [Salmonella enterica
 subsp. enterica serovar Typhi]"
 /codon_start=1
 /transl_table=11
 /product="putative heat shock protein"
 /protein_id="AAM83830.1"
 /db_xref="GI:21956912"
 /translation="MSEDMRYRVMSLCVTHSLTHNDRVHMINKPKQFHISLLHPRYW
 LWTGELFVLVLQVPLPLKNGLWGLGRTSMRFLKRRVSIARRNLELCPPDMQKIVF
 EQTIGNFESIGMLGTEGMWAFVSDTRIQRWFVSGLENLKRQQGKRGVLVIGVH
 MSLEIGRVMGQCPQMAMTYRPHNNKMYELVTVWGRMSNRKMLDRDLRDLVLRKQ
 GEAVWAFDQDYGPRGSVFAPLFAVEHAATTSGTFMRLARLAKPALLPLVLLRKKEGRG
 YDLLIQPALEDYPTIDDEATAASYNKVIKEMIRAPQYLWLHRRFKTRTIGEPSLY"
 /8768..10063
 /gene="y0237"
 /8768..10063
 /gene="y0237"
 /function="unknown"
 /note="residues 186 to 247 of 431 are 26.15 pct identical
 to residues 515 to 579 of 865 from GenPept :
 >dbj|BA873440.1| (AF003587) ORF_ID:all1741
 probable proteinase [Nostoc sp. PCC 7120]"
 /codon_start=1
 /transl_table=11

	Query Match	9.3%	Score 369;	DB 1;	Length 10201;
	Best Local Similarity	53.3%;	Pred. No. 3.3e-96;		
	Matches 814;	Conservative 0;	Mismatches 695;	Indels 18;	Gaps 1;
QY	858	CATGACAAATAACCAAAACCGTTATTGTTGTTGATCGGATCAGGCAGTAGTCCGCGCGCG	917		
Db	2372	CAAAACAATGGCAAGTTACTTTATCGCGGTAGAGCTAGGACCGGCAGTGCAGAGCTGG	2313		
QY	918	GATTTTGTGATCTCAACGGATCTGCTATCCCATGCCACAGAAAAAATCAGCACTACGG	977		
Db	2312	GGTATTGCAATTCAGGGCCGATGTTGTTGGCAGGCCGCGTGAATACCACTGTTTAA	2253		

Qy	978	GC	CAGCGGAAGCCGGTGGAAACAGATCCACGACGAGAGATCTGGCAGCGGGTGTGTTCA	TG	1037
Db	2252	ACCAAGGCTGATTTTGTGAGCAGCTCTCCGAGAAATATTTGGCAAGCGGTGTGTAATGC	2193		
Qy	1038	TATTCTGTAATCGGCTCACCTCGCAGACGTTTGTGCACAAAGTGTGGCAGGCATCGGTTT	1097		
Db	2192	GGTGGTGAATGCGGTTAATCAGCGCGATATCAACCCGATTCAGAGTCAAAAGGCGTGGGTTT	2133		
Qy	1098	TGATGCCACCTGTTCTCTGGTGTACTGGATAAAAACGGTGTATCCATTTCGCTGTTCAGGCC	1157		
Db	2132	TGATCGCACTTGTCTGCTGTGTGGATTAAGNAGGGAATCCACTACGCGTTAGGCC	2073		
Qy	1158	GGAAGGAGATGCAAGCAAAATATCATTTGTGGATGGATCACCGCGCCACCGAACAAAGC	1217		
Db	2072	TTACAGGCGCAATGAGCAAAATGTATCTTTGGATGGACCAACCGCGGATCACTCAGGC	2013		
Qy	1218	AGACCGAATCAATGCCACTCACCATCCGCTGCTGAACCTACGTCGGTGGTAAATTTCCGC	1277		
Db	2012	GGAGCGGATTAATGCCACCAAGCACCGCTCTCGAGTTTGTCTGGTGTGTGATTTCCGC	1953		
Qy	1278	TGAATTGGAAACACCGAAAATCTCTGGCTGAAAGAAAATATGCCAGAGATCTACGAACG	1397		
Db	1952	AGAAATCGACACCAAAAGTTGTTGTGTTAAAGCAACATATGCCCAACACGTGGAGTAA	1893		
Qy	1338	TGCGGGACAAATTTTTCGATCTGGCGGATTTTCTGACCTGGCGGGCTACCGGTGATTTAGC	1397		
Db	1892	TGTCGGCATCTGTTTGATCTACAGATTTTCTTACCTGGCGTGCCACGAAAGATGAAC	1833		
Qy	1398	CGGTTAGTATGCACTGTTTACCTGTAATAGCAGTGGCTGGCACATGATAAATTCGCTGGGA	1457		
Db	1832	CCGTTCTGCTGTCTCCACCGCTCTGCAAAATGGACCTATCTGGGCCATGAAGATCGCTGGGA	1773		
Qy	1458	TCAGATTAATTTCCGACCAATTTGGCCTTCGACAGTTTAGCGGATGAAGATTTTATTTCGCAT	1517		
Db	1772	TCCGTCCTACTTCAAAATGGTGGGGTGCCGATTTTACTGGATAACAACGGCGGCGAAGT	1713		
Qy	1518	TGCTCATCATATTTGTTTCTCCCGAAACACTTGTGGAATGTTTTAACAGCACAAGCCGC	1577		
Db	1712	CGGTGCGACGTTTAAACCCATGGGGCACCGTTTAGGCGATGGTTTTAAGTCACGCGGGCAGC	1653		
Qy	1578	GGCAGAGATGGGATTAATCCCGGCACACCTGTCTGCTGTAGGGTTTAATTTGATGCTCAGC	1637		
Db	1652	ATCAGAAATGGGTTGATCCCGGTCACGCGCTGAGTGTGTCGATTAATTGATGCTCAGC	1593		
Qy	1638	TGTTGGCATCGGTACGTTTGGCTAGAAGTGGAGCGCTGAACAAT-----	1683		
Db	1592	TGGACGATTTGGCATTTCTCGGTGCCAGTGGTGTGACGGGGGAAATGCTAATTGTATCG	1533		
Qy	1684	----CTCGCGTATGTTTTTCGGCACTTCTTCATGCACCACTGGCATCTACCACCTTCTCCCTC	1739		
Db	1532	CGGATTTGCGCTGATTTGGTGGCACCTCTACCGCCCATATGGCGATGTCCGCTCAGCAC	1473		
Qy	1740	GTTTGTACCGGTTGCTGGGGCCGTTAATTACAGTGGGATGGTTTCCAGGTCGTGTGGTTAGT	1799		
Db	1472	CTTTATCAGTGGTATTGGGGGCCATTAATTTCGGCCATTTTACCGGAATATTGGTGTGAA	1413		
Qy	1800	TGAAGCGGGCAAAAGTCTCGCGGACAGCTATTGAACAGCTACTTGTATTTCCATTCGCGC	1859		
Db	1412	CGAGGCGCGCAGTCGCAACGGGGCTTTGATTGACCATATTATTCAATCACATCTCTTG	1353		
Qy	1860	TGTTGAAGAAGCTCGGGAATGSCAACAGCTGTGAATCAGCCCTCCCGCTCGGCTTGC	1919		
Db	1352	CTATCCGGCACCTCTGGAACAGCTAAAAATAAGGGCGAAACCATCTATGAGGCGCTGAA	1293		
Qy	1920	TGATCGAATCCTCGAAAAACGGCGCAACCATCAGATGCTGTGCGCCTGGGCGAAAGGCT	1979		
Db	1292	TTATATTTTTCGTCAGATGGCGGTGAACCTGAAAATATTCGCTGTTCTCGACGAATGATAT	1233		
Qy	1980	ACAGTGGTCCGGAATTTCTCGGAATACGCGCCCTTTCGAGATCTCTCATGCCAGNC	2039		
Db	1232	TCATATGTTCCGCTACTTCACGGTTAACCGCTCTCCCGGTGCCAACCCCAATATTCAGCGG	1173		
Qy	2040	GGTAATTTTGGCCTGGGTATGAGCGAGATCTGGATTAATTTTACTTCGCTTGTATATCGC	2099		

Db 1172 CATATTACTGGCGCTGAACACTCTCTACACGCCAGAGATATGGCGCTCGGTATTTCGC 1113
QY 2100 TGGATTATCGGAATTTGGTTATGTCTGCGCCCAAAATTCGACGCTCAACACAGCGCAGGG 2159
Db 1112 AACCATTCAGCGCGCTGGCTGGCCACACGGCATATTATCGAAACCATCAATCAGATCG 1053
QY 2160 AGTAGTGAGTAANAATATCGTTATTAGCGGGGCTGCCGGCAGCATCCACTGTGACGACA 2219
Db 1052 CTATAACATTCGACACTATGATGGCCAGCGCGAGGACCAAGAACCCCATATTTGTACA 993
QY 2220 AATTCGTGGCAGATACCTGCGGTATTCGGTCAATACCGCATGCTGCGAACCGGTTTT 2279
Db 992 GGAGCATGGCATGCGACGGCTGCCCATGCTGTTGCCAGAGAGAGTGGCGCATGCT 933
QY 2280 ATTAGCTCGGGCATCTTCTGGTCTGCTGCCCGGAATATTGCACTCTGTTGGCGAAGC 2339
Db 932 GTTGGGAGTGCCATGATGGTACCGTCTGCTGGGGTGTGAGTCATTCGCCGAAGC 873
QY 2340 GATGCAACAATTCACCCATGTGGATAA 2366
Db 872 CATGGCGGCATGAGCCGCGATCGGTA 846

RESULT 10
AJ414158
LOCUS
DEFINITION
Yersinia pestis strain CO92 complete genome; segment 18/20.
ACCESSION
AJ414158 AL590842
VERSION
AJ414158.1 GI:15981524
KEYWORDS
Yersinia pestis.
SOURCE
Yersinia pestis
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.

REFERENCE
1 (bases 1 to 235050)
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G., Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L., Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M., Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G., Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Kariyasev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrall,B.G. Genome sequence of Yersinia pestis, the causative agent of plague. Nature 413 (6855), 523-527 (2001)
21470413
11586360

REFERENCE
2 (bases 1 to 235050)
Parkhill,J.
Direct Submission
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
Location/Qualifiers
1. .235050
/organism="Yersinia pestis"
/strain="CO92"
/db_xref="taxon:632"
/note="blovar: Orientalis"
48. .830
/gene="YPO3604"
48. .830
/gene="YPO3604"
/note="no significant database hits"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC93073.1"
/db_xref="GI:15981525"

FEATURES
source

gene
CDS

/db_xref="SPTREMBL:Q82B25"
/translation="MSDTPQHIIIKTGTDPNRPENFAIREINKINHPARPEVNMGL
IESALTLFRTHGVLDQFVAVYTLARTKNGLAGTCEGLAGVGVGQMDHLWPEQ
QARSILWFNTRVSNQLRQHDFTRDDRLVYRAERALLQVLYDKQLQKVELKVPRIEN
LLYLQMNTAKKLESASDAKQOATAPLKMPMVLVSLVPEAEPVRTAAAPPAANIE
VRSPRRLLNGRCGCGALVPVSCVYVWRLRFTLPSSNRCSK"
827. .1360
/gene="YPO3605"
827. .1360
/gene="YPO3605"
/note="Similar to Arabidopsis thaliana F15p11.1 protein
TR:Q9S9U2 (EMBL:AF160760) (187 aa) fasta scores: E():
0.024, 31.5% id in 143 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC93074.1"
/db_xref="GI:15981526"
/db_xref="SPTREMBL:Q82B24"
/translation="MTVLATOPESAALLWLNRPDVATYGEQLSTLENLSPLVLTAD
OSVAMARQWSPDSQVAESQWAEARLVEARIGLAGTDSYFQZQORLHALSKLLEQE
RSRSLTISYLKTAIVYQMTLNREIPLEELLRLQLAVSADHQSPASVLIKQIDDRWN
ALLSRHHLTQOTNSAR"
1382. 3784
/gene="YPO3606"
1382. 3784
/gene="YPO3606"
/note="Similar to Escherichia coli VgrE protein TR:052672
(EMBL:AF044503) (713 aa) fasta scores: E(): 0, 49.9% id in
579 aa, and to Escherichia coli VgrE protein TR:052660
(EMBL:AF044499) (702 aa) fasta scores: E(): 0, 49.3% id in
576 aa. Similar to YPO3613 (78.0% identity in 803 aa
overlap) and to YPO0764 (77.2% identity in 615 aa
overlap)."
/codon_start=1
/transl_table=11
/product="putative Rhs accessory genetic element"
/protein_id="CAC93075.1"
/db_xref="GI:15981527"
/db_xref="SPTREMBL:Q82B23"
/translation="MFAHDKANNKNAADKTGTVTPTHTVSPAAPTVSGAOMTEMAGG
IGDAVGTAGKRINQAAEVAKTALDAKAKVLDGVTPTNIVAGTAPTVSGAOMTEMA
GGIGDVGAGTAGKRINQAAEVAKTALDAKAKVLDGVTPTNIVAGTAPTVSGAOMTEMA
ISRLVKVSGGQFLLTASLPQTFAVDFTLSEMLSPFVLNGLASADPAVDFAAV
LDEDTALFIMREGVLQRTSGVAFSGEDGTGFHQTRYSMVIRPALWRTSLRNARIF
QOASVEEITITLLKENGIDFAFGFRHPHPVREFCVQESDFDIQRULTAEGLFYF
FEFSAGKNTVVADVVGSLPKGASLPYNVAAQAEISITTTTSAOVRPAEAGVQK
YTFKNPNAAFSEOSGELQNPDPYEHFDPGRKDAQHGODFTRYRLDALRNDANL
GCGASNDFTLQPLFSLYNHPRGDLNHAWQLLGHOSKQMOALEQSGDQGTVLFN
HFSFIPHTQWRPTPLAKPMDGQPIAMVVPPEEYICDEYGRQLQLWDRYQGSN
DNSSCWIRVTPWAGWGMLAIPRIGQEVVVDVFLHGDPDQPIVTRTHANNIPPS
LPASKTQAFRSKTHKGEGYNEMREFDAGKGLFMHAKQDMTAVLNDKRDVTDQDH
SEHIGDQSVTVRNQSNITQNDRREVTVDQQTVEGNDYQLVWKEKKEFTKIRYT
EVHEDETITVTKSIHAKQGDISTPNAGITITHDGAIVLQGVIRLAADMIDLNP
EE"
1421. 1441
/gene="YPO3606"
/note="PS00154 E1-E2 ATPases phosphorylation site."
3038. 3103
/gene="YPO3606"
/note="Predicted helix-turn-helix motif with score 1023
(+2.67 SD) at aa 553-574, sequence IRLQLWDRYQGSNDNSCWIR"
3790. 4248
/gene="YPO3607"
3790. 4248
/gene="YPO3607"
/note="Similar to Pseudomonas aeruginosa hypothetical
protein PA0094 TR:AG03484 (EMBL:AE004448) (144 aa) fasta
scores: E(): 8.9e-06, 24.8% id in 141 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC93076.1"

misc_feature

misc_feature

gene

CDS

```

/db_xref="GI:15981528"
/db_xref="SPTREMBL:O8ZB22"
/translation="MPENNINTYCIQEGTLECEPGEFTDRTVNLFMKCKPACAAALANIA
RDVPEAHEVLDYVQROIGILKNLKYKVKHQIDYGAQHIFGEVVEATYMAETRR
WQROQVAIQVGDHMIIFTATNASFFNPQOQATWEQIHSPAPWAREAGHV"
4241..7105
/gene="YPO3608"
CDS
4241..7105
/gene="YPO3608"
/notes="Similar to Escherichia coli Core protein TR:O52666
(EMBL:AF044501) (1420 aa) fasta scores: E(): 8.3e-32,
27.7% id in 824 aa, and to Escherichia coli RhsD protein
precursor SW:RhsD_ECOLI (P16919) (1426 aa) fasta scores:
E(): 1.4e-28, 27.3% id in 842 aa"
/codon_start=1
/transl_table=11
/product="putative exported protein"
/protein_id="CAC93077.1"
/db_xref="GI:15981529"
/db_xref="SPTREMBL:O8ZB21"
/translation="MFEARVDDKLVHSSALAGFIIGSIIGAAVIFAASAAVAAIVLT
GGATLVATGPIVGMVTTLGVAVGLIIRSVGEKIGSMCHDVQGITGSKNVKNSKR
AAHVELSTVACKDSDAIOIRMAEGSSNIFINSKAAVRLEKDTCTDAVVDSSNVTGG
GRVQYLDIKREISDEMARDLSEKLFIVAGLAGIFGAQKAGCFGLCKSLKIALGEMAG
AAAGYLEGQSGVIAIRFYGVYDVIVISQKLLTGGEDDTDFILPFIPLHWSRIYRSEN
HHVGAIGQGSLSWVRSRKRKEDDSIVYQNDREIVFPLIKRGERVSPFTEHWLART
ERTVATISSPFCTFIEAFSEAGVAKLASLEDLNGHALYFSYDDIGOLKIKISTGY
GVYCEKGRGLSVACVKGCTPCTLVRYOYNEOHLYSVTNREGQITROPQYHGLIN
KLADRGLECRYWADLGGTPRTHSNATNLGEQWQFDYDINDQOTLTLDTLNTQGTACW
GYNAQHLIDYDFDGKYAFDYNLMPVRVVLVAGERTLVLYDALRIQIDPLK
RETHIDYHRNSLRVRVYDPGVQWKEGYDRTGLKENAPDGCVTLIYHPGASSLPE
RITNAVGAOHLGWERHQLTHTDCSGKLTREYDIDGHLTVIDAENHSTHYSYNR
LGQPTGRYADGRKOLRYNAOGLVEQFDPQVROLHRYNLRGQPVSTFDRLQREVR
VRYDCHQMTLNDANGGOYHFRWSSGGLVVEQYDNLVRRYVYCESGMLALETTA
PTVDDLLYRQVDFDYDAGGEMTQRLTGMSTATRYDWDIMBRLLAERYPTAVGEQAGI
VHGVLVADKAGHLHLESGDLGAVTQMDLHLAALTLPDQGTLSWLYRAGHVSA
IRHGDTLISFSDNHLHREVSRTOGILTOYRDYDAMGRRL"
4298..4366
/gene="YPO3608"
/notes="one of 2 probable transmembrane helices predicted
for YPO3608 by TMHMM2.0"
4385..4453
/gene="YPO3608"
/notes="one of 2 probable transmembrane helices predicted
for YPO3608 by TMHMM2.0"
7133..8503
/gene="YPO3609"
7133..8503
/gene="YPO3609"
/notes="Similar to Escherichia coli hypothetical Core
protein RhsD TR:AAA24544 (EMBL:LI9084) (1426 aa) fasta
scores: E(): 4.6e-14, 28.3% id in 473 aa, and to
Escherichia coli RhsE protein SW:RhsE_ECOLI (P24211) (682
aa) fasta scores: E(): 1.6e-14, 26.8% id in 522 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC93078.1"
/db_xref="GI:15981530"
/db_xref="SPTREMBL:O8ZB20"
/translation="MAADLLPQGDIDWIRKFSFDIAGELSMATDFIRGEQYRYDADEGR
LTDNRERHQUSVAEDFYADNADNLLRLKLPFDYDLPYVANNRLTQWQHYREY
DANGNTRHAGRMQHFAYDDNLLRLKLPFDYDLPYVANNRLTQWQHYREY
RGAKEKTRQDFDFTQGLLQEQHTGNATYIYDPNESYTPPLARVDORHGETSOVY
FTFDINCTPLDVTGDEGKHKRWGSKYHAWGVTRQNVSDPQSVRPAOPLRYGVYS
DDEGLHYNTFRYIDPEIGRFSQDPIGLAGGINLYQYGNPLTWIDPWGAWFGVD
TGSFDLPVVGQSLNVEITMQARGKDFEAFKLAGISKADATGYTHHNLNDFDPSV
GKTMPLQVLTSAHEATFPHAGSVSQSEKHEFLPSQYSADAVAISHSKGWLKRIPK
ALRSGC"
8505..8990
/gene="YPO3610"
8505..8990
/gene="YPO3610"

```

```

/notes="Similar to Bacillus subtilis Yokk protein TR:O31996
(EMBL:D99115) (192 aa) fasta scores: E(): 0.043, 26.6% id
in 154 aa, and to Bacteriophage SPBc2 hypothetical protein
Yokk TR:O64025 (EMBL:AF020713) (192 aa) fasta scores: E():
0.043, 26.6% id in 154 aa"
/codon_start=1
/transl_table=11
Query Match 9.3% Score 369 DB 1 Length 235050;
Best Local Similarity 53.3% Pred. No. 4.8e-96;
Matches 814; Conservative 0; Mismatches 695; Indels 18; Gaps 1;
Qy 858 CATGACAAATAACCAAAACCGTTATTGTTGATGTGGGATCAGCAGTGTCCGCGCGG 917
Db 41868 CAAACAAATGCGCAAGTTACTTATCGGCGTAGACGTAGGACCGGCGTGCAGAGCTGG 41927
Qy 918 GATTTTGTATCTCAACGGATCTCTGCTATCCATGCCACACAAATAATACAGCTACGGG 977
Db 41928 GGTATTTCGATTTCAGGGCCGTATGTTGGCGAGCCCGCTGAATACCATCTTTAA 41987
Qy 978 GCGCAGCGGAAGCCGCGTGGGAACAGTCCAGCCAGGAGATCTGGCAGGCGTCTGTTTCATG 1037
Db 41988 ACCAAGCTCATTTTGTGAGCAGTCTTCAGAGAAATATTTGGCAAGCGGTGTGTAATGC 42047
Qy 1038 TATTGCTAATCGCTCACTCTGGCAGACGTTTGTGCACAAAGTGTGGCAGGCATCGGTTT 1097
Db 42048 GGTGCGTGATCGGTTAATCAGGCCGATATCAACCCGATTTCAGGTCAAAGGCTGGGTTT 42107
Qy 1098 TGATGCCACCTGTTCTCTGTTGGTACTCGATGATAAAGCGTATCCATTTGCTGTCAGCCC 1157
Db 42108 TGATGCGACTTGTCTGCTGTTGGTGTGATTAAGAAAGGGAATCCACTGACGGTTAGCCC 42167
Qy 1158 GGAAGGAGATCAAGCAAAATATCATTTGTGTGGATGATCACCGCGCACCAACAAGC 1217
Db 42168 TTCAGGGCGCAATGAGCAAAATGTGATTTGTTGGATGACCAACCGGCGATCAGTCAGC 42227
Qy 1218 AGAGCGAATCAATGCCACTCACCATCCGCTGCTGAACCTACGTCGGTGGTAAATTTCCGC 1277
Db 42228 GGAGCGGATTAATGACCAACAGCACCCGCTCTGGAGTTTGTGCGTGTGTTGATTTCGCC 42287
Qy 1278 TGAATGAAACACACCGAAATTCCTGCTGNAAGAAATATGCCAGAGATCTACGAACG 1337
Db 42288 AGAATGACAGACACCAAAAGTTGTTGTTAAAGCAACATATGCCGGAACACGTTGAGTAA 42347
Qy 1338 TCCCGGACAATTTTTCGATCTGGCGGATTTCTGACCTTGGCGGCTACCGTGTGTTAGC 1397
Db 42348 TGTGCGCATCTGTTTGTATCTACAGATTTCTTACCTGGCTGCCAGGAAGATGAAC 42407
Qy 1398 CGTTTCAGTATGCACTGTTTACCTGTAATAGGACGTGGCTGGCACATGAAATTCGCTGGGA 1457
Db 42408 CCGTTTCGCTCTGTTCCACCGCTGCAAAATGGACCTATCTGGGCCATGAAGATCGCTGGGA 42467
Qy 1458 TCCAGATTATTTCCGCACCATTCGCTTGCAGAGTTAGCGGATGAAGATTTTATTCGGAT 1517
Db 42468 TCCGTCTTACTTCAAAATTTGGTGGGCTGGCGGATTTACTGGATAACAACCGCGGGAAGAT 42527
Qy 1518 TGGTCATCATATTTTCTCCCGGAACACCTTGTGGAATGTTTAAACAGCACACACCGC 1577
Db 42528 CCGTTCGCGGCTTAAACCCATGGGGGACCGCTTAGGCGATGTTTAAAGTCACCGCGGACG 42587
Qy 1578 GCGAGATAGGATTAATCTCCCGGACACCTGCTGCTCTAGGTTAAATGATGCTCAGC 1637
Db 42588 ATCAAAATGGGTTGATCTCCCGGTACGCGGCTAGTGTGCGATTTATGATGCTCAGC 42647
Qy 1638 TGGTGGCATCGTACGGTTGGCTAGAGGTGGAGCGCTGAACAAT----- 1683
Db 42648 TGGGACGATTGGCATTTCTCGGTCGCGAGTGGTGTGACGGGGGAAATGCTAACTTTGATCG 42707
Qy 1684 -----CTCCCGTATGTTTTCGGCACCTTCTTCATGCAACCATGCACTACGCTTCCCTC 1739
Db 42708 CCGGATTCCGCTGATTGTTGGCGACCTCTACCGCCCATATGCGGATGTCGCGCTCAGCACA 42767
Qy 1740 GTTTGTACCGGGTGTCTGGGGGCGGTATTACAGTCGCGATGTTCCAGGTCGTGTTGTTAGT 1799

```

Db	42768	CTTTATCACTGGTATTGGGGGCCATATTATTTCGGGCCATTTTACCGGAATATTGGTTGAA	42827
Qy	1800	TGAAGCGGGCAAGTCTGCGGGAGCAGCTATTGACCACTACTTGAATTCATCCCGC	1859
Db	42828	CGAGCGGGCACTCGCAACGGGGCTTTGATGACCATATTATCAATCATCATCTTGG	42887
Qy	1860	TGTTGAAGAGCTCGGGAATAGCACAAGCTGTGAATACGCCCTTCCCGCTGCGCTTGC	1919
Db	42888	CTATCGGCACCTGCTGGAACAGCTAAAAATAAGGCGGAACCATATGAGGCGCTGAA	42947
Qy	1920	TGATCGAATCTCGAAAAACGGCCCAACCATCAGATGCTGCGCCCTGCGCAAGGGCT	1979
Db	42948	TTATATTTTTCGCTCAGATGCGGGGTGAACCTGAAATATTCGCTGACGAATGATAT	43007
Qy	1980	ACACGTGTGTCGGGAATTTCTCGGAATCGCGCCCTTCGCAGATCTCTCATGCCAGAGC	2039
Db	43008	TCATATGTTGCGCTACTTCCAGGTAACGCTCTCCCGTGCACACCAATCTGACGGG	43067
Qy	2040	GGTAATTTGCGCTGGGATGAGCGAGATCTGGATAATTTACTGCGCTTGTATATCCG	2099
Db	43068	CATTATTACTGCGCTGAAACTCTCTACCACGAGAGATATGGCGCTCGCTTATTGGC	43127
Qy	2100	TGATATTCGGAATTTGTTATGTTCTGCGCCCAATTTCTGAGCTCAACAGCGCAGGG	2159
Db	43128	AACCATTCAGGCGCTGGCTTGGGCACACGGCATATTTATCGAAACCATGAAATGG	43187
Qy	2160	AGTAGTGAGTAATAATCTGTTATTAGCGCGGTGCGCGGAGCATCTCACTGGTACGAA	2219
Db	43188	CTATAACATCGACATGATGCGCAGCGCGGAGGCGCAGACCAATTTGTTGACA	43247
Qy	2220	AAATCTGCGAGATACCTGCGGTATTTCGGTCAATTAACAGCAATGCTGCGAACCGT	2279
Db	43248	GGAGCATGCGAATGCGACGGCTGCGCCATGCTGTTGCGAGAGAGTGAAGCGATGCT	43307
Qy	2280	ATTAGCTCGGCATCTTGTGCTGCTGCGCGGAATATTTGACCTTCTGTTGGCGAAGC	2339
Db	43308	GTTGGGAGTGCGCATGATGGGTACCGTCTGCTGCTGGGTGTTGAGTCACTCCCGAAGC	43367
Qy	2340	GATGCAACAATTCACCATGTGGATAA	2366
Db	43368	CATGGCGGCATGAGCCGATCGGTAA	43394
RESULT 11			
BC014947			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
1 (bases 1 to 1923)			
Submitted (01-OCT-2001) National Institutes of Health, Mammalian			
Gene Collection (MGC), Cancer Genomics Office, National Cancer			
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
USA			
NIH-MGC Project URL: http://mgc.nci.nih.gov			
Contact: MGC help desk			
Email: cgapbs-remail.nih.gov			
Tissue procurement: ATCC/DCD/DTP			
CDNA Library Preparation: Rubin Laboratory			
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Genome Sequence Centre,			
BC Cancer Agency, Vancouver, BC, Canada			
info@bcgsc.bc.ca			
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,			
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,			

Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 34 Row: 1 Column: 7
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10433590.

FEATURES	source
Location/Qualifiers	1..1923
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="MGC:22958 IMAGE:4871664"	
/tissue_type="Skin, melanotic melanoma, high MDR."	
/clone_lib="NIH_MGC_49"	
/lab_host="DH10B-R"	
/note="Vector: pONB7"	
434..1753	
/codon_start=1	
/product="Unknown (protein for MGC:22958)"	
/protein_id="AAH14947.1"	
/db_xref="GI:15928959"	
/translation="MWLDHRAVSQVNRINETHKSHVQVGVMSVEMOAPKLLWKEN LEVICWDKAGHFDLPDLFSKATGVTARSLSLVCKWTYSAEKGDSDSPKMGLED EVAQYKIGNQVLPPLPGASLGNLTPEAROLDLPGTAVASLIDAIWAGLVGAD VRGHLICEGQVTVSLAVICGTSCHMGISKDPFVPGVWGPYSFVPGWLNREGG QSVTKLIDHWVQGHAAPELOVATKQCSYIAYLNHSLDLIKKAPGVFTVDLHV WPFHGRNSPLADLTAKMTGLKLSODLIDLTALYLAQIAIGRTFTIEMAEAG HSISFLFCGLSKNPLFVQMHADITGMVPLVSQEVSVLGAALGACASGDFASVQ EAMAKMSKVGKVVPRLODKKYYDKYQVFLKVEHQEYLAIMND"	
BASE COUNT	486 a 422 c 520 g 495 t
ORIGIN	

...	Query Match	7.6%	Score 302.4;	DB 9;	Length 1923;
	Best Local Similarity	52.1%	Pred. No. le-76;		
	Mismatches 827;	Conservative	0;	Mismatches 716;	Indels 45; Gaps 5;
Qy	879	TATTGGTGTGATGGGATCAGGCAGTGTCCGCGCGCGGATTTTGTATCTCAACGGATC	938		
Db	133	TCTGGGTGTGGAGTGTGGAACAGGCGAGTGTCCGTGCGAGCTTGTGGAGCAGAGTGGGT	192		
Qy	939	TCTGCTATCCCATGCCACAGAAAAATACGACTAGCGCGCGCAGCGGAGCGCGTGA	998		
Db	193	CCTGTTGGCTTTTGGCAGACCACCAATTAAGAAATGGGAGCCCCAGCTTCAACCACTGA	252		
Qy	999	ACAGTCCAGCCAGAGATCTGGCAGCGGTCTGTTCATGATATTCGTAATCGCGTCACTCT	1058		
Db	253	GCAGTCTCCGAGGACATCTGGCTGCGTGTGTGTGCACAAAGAAAGT-----TGT	306		
Qy	1059	GCACAGCTTTGTGCACAAAGTGTGGCAGGATCGGTTTGTATGCCACCTGTCTCTGT	1118		
Db	307	ACAAGGATTTGATTTAAACCAAAATTCGAGGAGCTTGGTTTGTATGCCACGTCTCTCTGT	366		
Qy	1119	GGTACTGGAATAAAACGGTGTATCCATTGCTGTACGCGCGGAGGAGATGCAAGCAAA	1178		
Db	367	TGTTTGGATAGCAGTTTCAACCCATTACAGTCAACAGGAGGGGATTCCTCATCGAA	426		
Qy	1179	TATCATTTGTGGATGGATCACCAGCCACCGAACAGAGCGGAATCAATGCCACTCA	1238		
Db	427	CGTCATCATGTGGCTGGACCATCGAGCAGTCAAGTAACTAAGAGTCAATGAGACCAA	486		
Qy	1239	CCATCGGCTGTGAACACTAGCTCGGTGGTAAAAATTTCCGCTTGAATGGAACACCGGAAAT	1298		
Db	487	GCACAGTGTCTCCACAGTACGTGCGGGGGGTGATGTCTGTGGAAATTCAGCGCCGAACT	546		
Qy	1299	TCTTGGCTGAAGAAATATCCAGAGAT---CTACGAACGTCGCGGAGCAATTTTCGA	1355		
Db	547	TCTTGGCTGAAGAGAACTTTCAGAGAGATTTGCTGGGATAAGCGGGACATTTCTTTGA	606		

QY 1356 TCTGCGCGATTTCTGAACCTGCGGCGCTACCGGTGATTTAGCGCGTTCAGTATGCACGT 1415
Db 607 TCTCCGGGACTTCTTATCGTGAAGCAACAGGTGTACAGCACGGTCTCTCTGCTCCCT 666
QY 1416 TACCTGTAATGGACGTGCTGGCAGATGAATAATCGCTGGGATCCAGATTAATTTCCGAC 1475
Db 667 GTGTGTAAGTGGACATATTCAGCA---GAGAAAGCTGGGAGCAGAGTTTCTGGAAAT 723
QY 1476 CATTTGGCCCTTCGACAGTTAGCGGATGAAGATTTTATTCGATTTGGTCATCATATTTGTTTC 1535
Db 724 GATTGGTTTGAAGACTTTGTTGCAGATAATACAGCAAAATAGGAAACCAAGTCTTACC 783
QY 1536 TCCCGGAACACCTTTGGAATATGTTTACAGCAACAGCGCGGAGAGATGGGATTAAT 1595
Db 784 TCCCTGGATTTGGTGGCGAGCTTCACTATTGATGCCATGCGAGGAGTCTAGGAGTGAT 903
QY 1596 CCCGSCACACCTGCGCTGAGGTAAATGATGCTCAGCTGCTGGGATCGGTACGCT 1655
Db 844 CCTTGGATTTGGTGGCGAGCTTCACTATTGATGCCATGCGAGGAGTCTAGGAGTGAT 903
QY 1656 TGGCTGAGAGGTGGAGCGC-----TGAACAATCT 1685
Db 904 TGGGCGAGATGTAGAGGCGACGGCTCATCTGTGAGGGCAGCCAGTCACTGACGCT 963
QY 1686 CGCGTATGTTTTCGGACCTTCTTCATGACCAATGGATACACTTCTCCCTGCTTTGT 1745
Db 964 GGCTGTCATCTGTGGAACGCTTCTTGTGCATGGGATCAGCAAAAGACCGATTTTGT 1023
QY 1746 ACCGGTGTCTGGGGCGTATTACAGTCCGATGTTCCAGCTCTGCTGTTAGTTGAAG 1805
Db 1024 ACAGCGCTCTGGGGCGTATTCTTCAGCCATGGTACCTGGGTCTGCTGCTGAATGAAG 1083
QY 1806 CGGCAAGTGTCTGGGAGCAGCTATTGACAGCTTACTTGTATTCATCCGCGCTTTCA 1865
Db 1084 TGGTCAGACGCTTACTGGAATATTGATAGCACCATGGTACAGGCCATGCTGCTTTCC 1143
QY 1866 AGAAGCTCCGAAATGGCAACAGTGTGAATCAGCCCTTCCCGCTGCTGCTGCTGATCG 1925
Db 1144 AGAATACAAAGTAAAGGCGACAGATGCCAGAGTATATATGCAAT---ATTGAAACAG 1200
QY 1926 AATCTCGAAAAACGCGCAACCATCAGATGCTGCGCCCTGGCGAAAGGCTACACGT 1985
Db 1201 TCACCTGGATCTGTTAAGAAGCTCAGCTGTGGTTCCTTACTGTTGATTTACATGT 1260
QY 1986 GGTGCGGAAATTTCTCGAAATCGCGCCCTTTCGACATCTCTATGCCAGAGCGGTAAT 2045
Db 1261 TTGGCCAGATTTCCATGGCAACCGGTCTCCCTTAGCAGATCTGACACTAAAGGCGATGT 1320
QY 2046 TTGTGCGCTGGGTATGGAGCAGATCTGGATAATTTACTGCGCTGTATATCGCTGGATT 2105
Db 1321 CACCGGATTTGAACCTGTCTCAGGACCTTGATGATCTTGCCATTTCTACCTGGCCACAGT 1380
QY 2106 ATGCGGAATTTGTTATGCTGCGCAAAATTTCTCAGCGCTCAACAGCCAGGGAGTAGT 2165
Db 1381 TCAAGCCATTTGTTGGGACCTCGCTTTCATTATAGAAGCCATGGAGGCGAGGCGACATC 1440
QY 2166 GAGTAAAAATATCGTTATTAGCGGCGTCCGCGGAGCATCTCAGTGGTACGACAAATCT 2225
Db 1441 ATCAGTACTCTTTTCCCTATGTGGAGGCTCAGCAGAAATCCCTTTTGTGCAATGCA 1500
QY 2226 GGCAGATACCTCGGTATTCGGTTCATTACACGCAATGCTGCGAACCCGTTTATTTAG 2285
Db 1501 TCGGACATTACTGGCATGCTGTGCTCTGTCGCAAGAGTGGAGTCCGTTCTTGTGGG 1560
QY 2286 CTCGGCATTTCTTGTGCTGTCGCGGAAATATTGCAACCTTCTGTTGGCGAAGCGATGCA 2345
Db 1561 TCGTGTGTTCTGGGTGCTGTGCTCTCAGGGATTTCTGCTGTACAGGAAGCAATGCG 1620
QY 2346 ACAATTCACCATGTTGGATAAATATATTATTCGCAAGAACCTATCATCTCTTCATCA 2405
Db 1621 AAAATGAGCAAGTTGGGAAGTTGTGTTCCGAGACTACAGGATAAAAAATACTATGA 1680
QY 2406 TCGTCGATATGAGGCTTATAAGCAGTTG 2433

Db 1681 TAAGAAATACCAAGTATTCTCTGAAGCTG 1708
RESULT 12
AX427495
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
REFERENCE
AUTHORS
Hodgson, D.M., Lincoln, S.E., Russo, F.D., Spiro, P.A., Banville, S.C.,
Bratcher, Shawn, R., Dufour, G.F., Cohen, H.J., Rosen, B.H., Shan, P.,
Chalup, M.S., Hillman, J.L., Jones, A.L., Yu, J.Y., Greenawalt, L.B.,
Panzer, S.R., Roseberry, A.M., Wright, R.J., Chen, W., Liu, T.,
Yap, P.E., Stockreher, T.K., Amshy, S. and Fong, W.T.
Molecules for diagnostics and therapeutics
Patent: WO 0121836-A 5 29-MAR-2001;
Incyte Genomics, Inc. (US)
Location/Qualifiers
1..1991
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 053778.11.dec"
BASE COUNT 468 a 467 c 534 g 516 t 6 others
ORIGIN
Query Match 7.6%; Score 300.8; DB 6; Length 1991;
Best Local Similarity 52.0%; Pred. No. 3e-76;
Matches 826; Conservative 0; Mismatches 717; Indels 45; Gaps 5;
QY 879 TATTTGGTGTGATGTTGGGATCAGGAGTGTCCGCGCGGATTTTGTATCTCAACGATC 938
Db 227 TGTGGGTGTGGAGCTTTGGAACAGGAGTGTCCGTGTGAGCTCTGTGGCAGCAGAGTGGGT 286
QY 939 TCTGCTATCCATGCCACAGAAAAATCAGCTACGCGCGCAGCGGAGCGCGTGA 998
Db 287 CTGTTGGCTTTTTCAGAGCAGCCAAATTAAGAAATGGAGCCCGCCAGTTCAACCCATGA 346
QY 999 ACAGTCCAGCCAGGAGATCTGGCAGCGGTCTGTTTCATGTTTCGTAATCGCTCACTCT 1058
Db 347 GCAGTCTCCGAGGACATCTGGCTGCTGCTGTTGTCACAAAGAAAGT-----TGT 400
QY 1059 GGCAGAGTGTGTGCAAAAGTGTGGCAGGATCGGTTTTGTATGCGCACCTGTTCTCTGT 1118
Db 401 ACAAGGATTTGATTTAAACCAAAATTCGAGGACTTGGGTTTGTATGCGACGTTCTCTGT 460
QY 1119 GGTACTGGATAAAACGGTGCATTCCTGCTCAGCCCGGAGGAGATGCAAGCAAAA 1178
Db 461 TGTGTTGATAAGCAGTTTCAACCCATTACCAAGTCAACAGTAAAGGAGTTCCCATCGAA 520
QY 1179 TATCATTTGTGGATGATCACCGCCACCGAAAGCAGAGCGGATTAATGCCACTCA 1238
Db 521 CGTCATCATGTGGCTGGACCATCGAGCAGTCAAGTTAAAGAGTCAATGAGACCAA 580
QY 1239 CCATCCGGTCTGAACTAGCTCGGTGGTAAAAATTTCCGCTGAAATGAAACACCGAAAT 1298
Db 581 GCACAGTGTCCCTCCAGTACGTCCGGGGGTGTGATGCTGTGGAATATGCAGGCCCGAACT 640
QY 1299 TCTCTGGCTGAAGAAAAATATGCCAGAGAT---CTAGCAACGTGCGGAGCAAAATTTTCGA 1355
Db 641 TCTGTGGCTGAAGAGAACTTGAGAGAGATTTGCTGGGATTAAGCGGGGACATTTCTTTGA 700
QY 1356 TCTGGCCGATTTTCTACCTGGCGGTACCGGTGATTTAGCCGTTTCACTATGCACTGT 1415
Db 701 TCTCCGGGACTTCTTATCTGTGGAAGCAACAGGTGTACAGCAGCGGTCTCTCTGCTCCCT 760
QY 1416 TACCTGTAATGGACGTGGCTGGCAGATGAAAAATCGCTGGGATCCAGATTTATTTCGCGAC 1475

Db 761 GGTGTGTAAGTGACATATTTCAGCA---GAGAAAGGCTGGGACGACAGTTTCTCGGAAAT 817
QY 1476 CATTGGCCTTGCAGAGTAGCGGATGAAGATTTTATTCGATTTGGTCATCATATTGTTTC 1535
Db 818 GATTGGTTTGAAGACTTTGTCAGATATTTACGAAATAGGAACCAAGTGCTACC 877
QY 1536 TCCCGGAACACCTTGTGGAATGTTTAAACAGACACAGCGCGGCGAGAGATGGATTACT 1595
Db 878 TCCTGGAGCTTCTCTTGGAAATGGCTCACACAGAGGAGCAAGAGACCTTGGCCCTCT 937
QY 1596 CCGCGGACACCTGTGCTGTAGGGTTAATTGATGCTCAGCTGTGGCATCGGTACGGT 1655
Db 938 CCCTGGGATTTGGCGTGCAGCTTCACTCAITGATGCCATGCGAGGAGACTAGGAGTGAT 997
QY 1656 TGGCTAGAGAGTGGAGCGG-----TGAAACATCT 1685
Db 998 TGGGGCAGATGTAGAGGCGCGGCTCATCTGTGAGGGCAGCCAGTAGGCTACGGGT 1057
QY 1686 CGCGTATGTTTTCGGCACCTTCTTCATGCACCATGGCATCTTACCACCTTCTCCCTCGTTTGT 1745
Db 1058 GGCTGTCTGTGGAAGCTTCTTGTGCATGCGGATCAGCAAGACCGATTCTTGT 1117
QY 1746 ACCGGTGTCTGGGGCGGTATTACAGTGGATGGTTCAGGCTCTGTGTTAGTGAAG 1805
Db 1118 ACCAGCGCTCTGGGGCGCTTATTCTCAGCCATGGTACCTGGTCTGCTGCTGAATGAAG 1177
QY 1806 CGGGCAAGTGTCTGGGAGCAGCTATTGACACAGTACTTGAATTCCTCCGCTGTGTA 1865
Db 1178 TGGTCAGAGCTTACTTGGAAATTTAGACACATGTTACAGGCGCATGCTCTTTTC 1237
QY 1866 AGAAGTCTCGGAATAGGACACAGTGTGAATCAGCCCTCCCGCTGTGGCTGTGCTGATCG 1925
Db 1238 AGAAGTCAAGTAAAGGCCACAGCCAGATGCCAGATATATGATCAT---ATTGAAACAG 1294
QY 1926 AATCTCTGAAAACAGGCCAACCATCAGATGCTGTGCGCTCGCGGAAAGGGCTACAGT 1985
Db 1295 TCACCTGGATCTGATTAAAGAGGCTCAGCGCTGGGTTCCTGTTACTGTGATTAGATGT 1354
QY 1986 GGTGCGGGAATTTCTCGGAATCGCGCCCTTCGCGAGATCCATCGCCAGAGCGGTAA 2045
Db 1355 TTGCGCAGATTTCCATGCGCAACGGTCTCCCTTAGCAGATCTCACACTAAGGGCATGGT 1414
QY 2046 TTGTGCGCTGGGTATGAGGAGATCTGGAATATTTACTGCGCTTGTATATCCTGGATT 2105
Db 1415 CACCGGATTTGAAACTGTCTCAGGACCTTGATGATCTTGCCATCTCTACCTGGCCACAGT 1474
QY 2106 ATCGGGAATTTGTTATGGTCTGCGCAAAATCTCGAGCTCAACAGCGGAGGAGTAGT 2165
Db 1475 TCAGGCATTTGCTTGGGACTCGCTTCATATAGAACCATGGAGCAGCGGCACTC 1534
QY 2166 GAGTAAATATCTGTTATTAGCGCGGTGCGGGCAGCATCTGTTAGCAGCAAAATCT 2225
Db 1535 AATCAGTACTCTTTCTCTATGTGGGCTTCAGCAAGATCCCTTTTGTGCAATGCA 1594
QY 2226 GCGAGTACTCTGGGTATTCCTGATATACCAGCAATGCTGCAACCGGTTTATAGG 2285
Db 1595 TCGCGACATTTACTGGATGCTGTGTCTGTGTCGCAAGAGGTGGAGTCCCTTCTTGGG 1654
QY 2286 CTCGGCATTTCTGTTGCTGCTGCGCGGAATATTGACACCTTCTGTTGGCGAAGCATGCA 2345
Db 1655 TGTGCTGTTCTGGTGGCTGCTGCGCTCAGGGGATTTGCGTTCTGTACAGGAACATGGC 1714
QY 2346 ACAATTCACCATCTGGATAAATATTATTATCCGCAAGACGCTATCAATCTCTTCATCA 2405
Db 1715 AAAAAATGAGCAAGTTGGGAAAGTTGTGTTCCCGAGACTACAGGATAAAAAAATACTATGA 1774
QY 2406 TCGTCGATATGAGGCTATATAGCAGTTG 2433
Db 1775 TAAGAAATACCAAGTATTCTCTGAAGCTG 1802

RESULT 13

AK090568
LOCUS
DEFINITION
Homo sapiens CDNA FLJ33249 fis, clone ASTRO2005081, linear PRI 15-JUL-2002
to L-RIBULOKINASE (EC 2.7.1.16).
AK090568
ACCESSION
AK090568.1 GI:21748754
VERSION
oligo capping; fis (full insert sequence).
KEYWORDS
Homo sapiens normal astrocytes (NHA5732) cDNA to mRNA,
SOURCE
clone lib:ASTRO2 clone:ASTRO2005081.
ORGANISM
Homo Sapiens
REFERENCE
1
AUTHORS
Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Kawao,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K.
and Isogai,T.
NEDO human cDNA sequencing project
unpublished
2 (bases 1 to 2485)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
Location/Qualifiers
1..2485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ASTRO2005081"
/cell_type="normal astrocytes (NHA5732)"
/clone_lib="ASTRO2"
/note="cloning vector: pME18SFL3-primary culture, normal
astrocytes"
BASE COUNT 618 a 550 c 605 g 712 t
ORIGIN
Query Match 7.1%; Score 282.6; DB 9; Length 2485;
Best Local Similarity 52.3%; Pred. No. 6.8e-71;
Matches 776; Conservative 0; Mismatches 664; Indels 45; Gaps 5;
QY 879 TATTGGTGTGATGTGGGATCAGGAGTGTCCGCGCGGATTTTGTGATCTCAACGGATC 938
Db 206 TGTGGGTGGAGCTTGGACAGCAGTGTCCGTGCGAGCTCTGTGGTGACAGGGTGGGT 265
QY 939 TCTGCTATCCATGCCACAGAAAAATACAGACTACGCGCGCAGCGGAGAGCGCGTGA 998
Db 266 CCTGTTGGCTTTTGCAGACCAGCAATTAAGAAGTGGAGCCCCAGTTCAACCACTATGA 325
QY 999 ACAGTCCAGCCAGAGATCTGGCAGCGGTCTGTTCTATGATTTCGTAAATCGGCTCACTCT 1058
Db 326 GCAGTCTCCGAGACATCTGGGCTGCGTCTGTGTTGTCACAAAGAAAGT-----TGT 379
QY 1059 GGCAGAGCTTTGTGCAAAAGTGTGGCAGGCATCGGTTTGTGATGCCACCTGTTCTCTGGT 1118
Db 380 ACAAGGATTTGATTTAAACCAAAATTCAGGACTTGGTTTGTATGCCACGTGTTCTCTGGT 439
QY 1119 GGTACTGGATAAAACGGTGTATCCATTTGCTGTAGCCCGCGGAGAGATCCAAAGCAAAA 1178
Db 440 TGTGTTGGATAAGCAGTTTCAACCATTTACCACTCAACAGAGAGGGGATTTCCCATCGAAA 499
QY 1179 TATCATTTGTGGATGATCACCGCGCCAGCAAGAGAGAGATCAATGCCACTCA 1238

```

      ||||| ||||| ||||| || || || || ||||| || ||||| || ||||| || ||
Db 500 CGTCATCATGTGGGTGGACCATCAGCAGTCAAGTAAACAGGATCAATAGACCAA 559
      || || || || || || || || || || || || || || || || || || || ||
Qy 1239 CATCCCGGTGCTGAACACGTCGCTGGTGAATAATTCGCTCAAAATGGAACACCGAAAT 1298
      || || || || || || || || || || || || || || || || || || || ||
Db 560 GCACAGTGTCTCCAGTACGTCTGGGGGGTGTGCTGTGGAATGACGGCCCCGAACT 619
      || || || || || || || || || || || || || || || || || || || ||
Qy 1299 TCTCTGGCTGAAAGAAATATGCCAGAGATCTAC---GAACGTCCCGGACAAATTTTTCGA 1355
      ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 620 TCTGTGGCTGAAAGAAACTTTGAGAGAGATTTGCTGGGATAAAGCGGGACATTTCTTTGA 679
      || || || || || || || || || || || || || || || || || || || ||
Qy 1356 TCTGGCCGATTTTCTGACCTGGCGGGTACCAGGTGATTTAGCGGTTCTAGTATGCACGT 1415
      || || || || || || || || || || || || || || || || || || || ||
Db 680 TCTCCACAGCTTCTATGCTGGAAGGCAACAGGTGTACAGCAGCGGTCTCTGCTCCCT 739
      || || || || || || || || || || || || || || || || || || || ||
Qy 1416 TACCTGTAAATAGCAGTGGCTGGGCATGAAATCGCTGGGATCCAGATTTATTTCCGCAC 1475
      ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 740 GGTGTGAAGTGACATATTCAGCA---GAGAAAGCTGGGACGACAGTTTCTGGAAT 796
      || || || || || || || || || || || || || || || || || || || ||
Qy 1476 CATTTGGCTTCAGAGTTAGCGGATGAAGATTTTATTCGCAATTTGGTATCATATTTGTTTC 1535
      ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 797 GATTGGTTTGAAGACTTTGTTGCAGATAATTACAGCAAAATAGGAAACCAAGTCTACC 856
      || || || || || || || || || || || || || || || || || || || ||
Qy 1536 TCCCGGAACACTTGTGGAATGTTTAAACAGACCAAGCCCGGACAGATGGATTTACT 1595
      ||||| ||||| ||||| || || || || || || || || || || || || || || ||
Db 857 TCTTGGAGCTTCTTGTGGAATGGCTCACACAGAGCGCAAGAGACCTTGGCCTTCT 916
      || || || || || || || || || || || || || || || || || || || ||
Qy 1596 CCCCGGCACACTGTCGCTGTAGGGTTAATTGATGCTCACCTGCTGTCATCGGTACGGT 1655
      || || || || || || || || || || || || || || || || || || || ||
Db 917 CCCTGGGATTCGGTTCGCGAGCTTCACTCATTTGATGCCCATCAGGAGACHTAGGATGT 976
      || || || || || || || || || || || || || || || || || || || ||
Qy 1656 TGGCGTAGAAGGTGGAGCGC-----TGAACAATCT 1685
      || || || || || || || || || || || || || || || || || || || ||
Db 977 TGGGCGACATGTGAGGGGCACGGCTCATCTGTGAGGGGCAGCCAGTCACTACGGCT 1036
      || || || || || || || || || || || || || || || || || || || ||
Qy 1686 CGCGTATGTTTTGGCAGCTTCTTATGCAACCATGGCATCTACCAGTCTCCCTGTTTGT 1745
      || || || || || || || || || || || || || || || || || || || ||
Db 1037 GGCTGTCTATCTGTGGAAGCTTCTTGTGCATGGGGATCAGCAAGACCCGATTTTGT 1096
      || || || || || || || || || || || || || || || || || || || ||
Qy 1746 ACCGGTGTCTGGGGCGGTATTACAGTGGCATGGTCTCCAGCTCTGTGTTAGTTGAAGG 1805
      || || || || || || || || || || || || || || || || || || || ||
Db 1097 ACCAGGCTGTGGGGCGCTTATTTCTAGCCATGGTACTGGTCTGGCTGGAATGAAGG 1156
      || || || || || || || || || || || || || || || || || || || ||
Qy 1806 CGGCAAGTGTCTGGGAGCAGCTATTACAGCAGTACTTATTCATCCAGCGCTGTGTA 1865
      || || || || || || || || || || || || || || || || || || || ||
Db 1157 TGGTCAGAGCCTTACTGGAAAATTTGATAGACCATGTGTACAAGGCCATGTGCTTTTCC 1216
      || || || || || || || || || || || || || || || || || || || ||
Qy 1866 AGAAGCTCGCAATGGCAACAGTGTGAATCAGCCCTCCCGTCTGGCTGTGTGATCG 1925
      || || || || || || || || || || || || || || || || || || || ||
Db 1217 AGAATAAAGTAAAGGCCACAGCCAGATGCCAGAGTATATATGATATTTGAACAGTCA 1276
      || || || || || || || || || || || || || || || || || || || ||
Qy 1926 AATCCTCGAAAAAAGCGCGCAACCATCATGATGCTGTGCCCTGGCGAAAGGCTACAGCT 1985
      || || || || || || || || || || || || || || || || || || || ||
Db 1277 ---CCTGGATCTGATTAAGAGGCTCAGCCTGTGGTTCCTTACTGTTGATTTACATGT 1333
      || || || || || || || || || || || || || || || || || || || ||
Qy 1986 GGTCCCGGAATTTCTCGGAAATCGCGCCCTTCGCGAGATCTCATGCCAGAGCGGTAAT 2045
      || || || || || || || || || || || || || || || || || || || ||
Db 1334 TTGCCAGATTTCCATGGCAACCGCTCTCCCTTAGCAGATCTGACGCTAAAGGCGATGTT 1393
      || || || || || || || || || || || || || || || || || || || ||
Qy 2046 TTGTGGCTGGGTATGGAGCGAGATCTGGATAAATTTACTCCCTTGTATATCGTGGATT 2105
      || || || || || || || || || || || || || || || || || || || ||
Db 1394 CACCGGATTTGAACATGTCTCAGGACCTTGTATGATCTTGCCATTCTCTACCTGGCCACAGT 1453
      || || || || || || || || || || || || || || || || || || || ||
Qy 2106 ATCGGGAATTTGTTATGTTCTGCCCAATTTCTGACGCTCAACAGCGGAGGAGTAGT 2165
      || || || || || || || || || || || || || || || || || || || ||
Db 1454 TCAAGCCATTCCTTTGGGACTCCTCTCATTTATAGGCCATGGAGCGAGCGGCGACT 1513
      || || || || || || || || || || || || || || || || || || || ||
Qy 2166 GAGTAAAAAATATCGTTATTAGCGCGGTGGCGGACGATCCACTGGTGTACGACAAATTTCT 2225
      || || || || || || || || || || || || || || || || || || || ||
Db 1514 AATCAGTACTCTTTTCTCTATGTGGAGGCTCAGCAAGATCCCTTTTGTGCAATGCA 1573
      || || || || || || || || || || || || || || || || || || || ||
Qy 2226 GGCAGATACCTTGGCGGTATTCGGGTCAATTACACCGCAATGTCTGCGAAGCGGTTTTATTAGG 2285
      || || || || || || || || || || || || || || || || || || || ||
```

```

Db 1574 TCGGACATTACTGCATGCCTGTGGGCTGTGCGAAGAGTGGAGTCCGTTCTTTGTGGG 1633
Qy 2286 CTCGGCCATTCTGTGCTGCGCGGAAATATTGCACCTTCTGT 2330
      || || || || || || || || || || || || || || || || || || || ||
Db 1634 TGCTGCTTCTGTGCTGCTGCTCAGGGATTTCCGTTCTGT 1678

RESULT 14
BC031708                               1842 bp      mRNA      linear      ROD 07-AUG-2002
LOCUS                               Mus musculus, Similar to RIKEN cDNA 2310009E04 gene, clone
DEFINITION                               MGC:25851 IMAGE:4194716, mRNA, complete cds.
ACCESSION                               BC031708
VERSION                               BC031708.1 GI:21618803
KEYWORDS                               MGC.
SOURCE                               house mouse.
ORGANISM                               Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE                               Strausberg, R.
TITLE                               Direct Submission
JOURNAL                               Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK                               NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT                               Contact: MGC help desk
                               Email: cgapbs-remail.nih.gov
                               Tissue Procurement: Jeffrey E. Green, M.D.
                               cDNA Library Preparation: Life Technologies, Inc.
                               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                               DNA Sequencing by: Baylor College of Medicine Human Genome
                               Sequencing Center
                               Center code: BCM-HGSC
                               Web site: http://www.hgsc.bcm.tmc.edu/cdna/
                               Contact: amg@bcm.tmc.edu
                               Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
                               Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
                               Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 30 Row: 0 Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.

FEATURES             location/Qualifiers
     source
     1..1842
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /map="FVB/N"
        /clone="MGC:25851 IMAGE:4194716"
        /tissue_type="Liver, normal. 5 month old male mouse."
        /clone_lib="NCI-CGAP_L19"
        /lab_host="DH10B"
        /note="Vector: pCMV-SPORT6"
     65..1228
        /codon_start=1
        /product="Similar to RIKEN cDNA 2310009E04 gene"
        /protein_id="AAH31708.1"
        /db_xref="GI:21618804"
        /translation="MMSGRDQEPSRYVGVGIDVGTGVSRAALVDQRLLLAFAPQPIKK
WEPOHNHFEOSSEDIAWACCLVTEKVVQIDAHRIGLGFDATCSLVVLDKERHPLPV
NHGCDSSRNVIWMLDHRACVQVHRINETHKRVLYVGVGSVEMQAPKLLWLENLRE
ICWDKAGFFDPLFLSKNATGTARSLCSLVCKRWITSAEKGWDSFKWMTGLDLDLD
DNTSKIGNLVLLPAAALIGLTTPAARELGLPSGIAVAASLIDAHAGGLGVIGADVRG
HGTTCGQPTVSLAVICGTSSCHGISKDPFVPGVWGPYYSANVPGFWLNEGQSV
TGKLIDHWQGHAPFELQAKATARNLQKHGIGHDTPGIKAYELEVTSSKAF"

BASE COUNT             507 a      387 c      453 g      495 t
ORIGIN

Query Match             5.7%; Score 226; DB 10; Length 1842;
```

Best Local Similarity 54.5%; Pred. No. 2.8e-54;
Matches 566; Conservative 0; Mismatches 430; Indels 42; Gaps 4;

QY	879	TATTGGTGTGGATCAGGAGTCTCGCGCCGGATTTTGTATCTCAACGGATC	938
Db	103	TGTGGCATCGATGTGGGAACAGGTAGTGTCCGTGCGCTCTAGTGGACAGAGAGGCT	162
QY	939	TCTGTATCCCATGCCACAGAAAAATCACGACTACGGGGCCAGCGGAAGCCGGTGA	998
Db	163	CTTTTGGCTTCGCGAGAGCACCCTTAAGAAGTGGGAGCCTCAATCAACCCACCA	222
QY	999	ACAGTCCAGCCAGAGATCTGCGCAGGGTCTGTTCATGTATTCGTATATGCGCTCACTCT	1058
Db	223	GCAGTCTCAGAGATATCTGGCGGCATGCTGCC-----TTGTCAAAAGGAAGTTGT	276
QY	1059	GGCAGAGTTTCTGCACAAAGTGTGCGAGGCATCGSTTTTGATGCCACTGTTCTCTG	1118
Db	277	TCAGGGATGTATGCATCGAATCCGAGGACTTGGTTTGTATGCCAGTGTCTCTG	336
QY	1119	GGTACTGGATAAAACGGTGTATTCATTTGCCCTGCAGCCCGGAAGAGATGCAAAACAAA	1178
Db	337	TGCTCTGGATAAAGAGTTCCATCTTTTACCAGTAAACCATGAAGGGACTCTCTCGAA	396
QY	1179	TATCATTTGTGGATGCATCCCGCGCCACCAACAGCAGAGCAATCAATGCCACTCA	1238
Db	397	CGTTATCATGTGGCTGGACCAACCGGGCTGTACGCCAGSTCCACAGGATTAATGAGACCA	456
QY	1239	CCATCGGTGCTGAATACGTACGTGGTGAATTTTCGCTGAAATGGAAACACCGAAAT	1298
Db	457	GCACAGAGTCTTCATGTATGTGGTGGTGTGATGTCTGTGGAGATGCAGGCTCCAAGCT	516
QY	1299	TCTCTGGCTGAAGAAAATATGCCAGAGAT---CTACGAACTGCGCGGACAAATTTTGA	1355
Db	517	CTGTGGCTGAAGGAGAACTTCAGAGAGATTTGCTGGGATAGGCGGACATTTCTTGA	576
QY	1356	TCTGGCCGATTTTCTGACCTGCGGGCTACCGGTGATTTAGCCGTTTCAGTATGCACTGT	1415
Db	577	TCTCCAGACTTCTTATCATGTAAGGCAACAGGTGTCAACAGCAGCTCTCTGTCTCTT	636
QY	1416	TACCTGTAATAGAGTGTGCTGGCATGAAATCGCTGGGATCCAGATTAATTTCCGCAC	1475
Db	637	GGTGTGCAATGCATACTACGCA---GAGAAAGTTGGGATGACAGCTTCTGGAAGAT	693
QY	1476	CATTGGCTTCAGAGTATAGCGGATGAAGATTTTATTCGATTTGTCATATATTGTTTC	1535
Db	694	GATTTGACTTGAAGATCTCATTTGATTAATACAGCAAAATAGAAACCTCGTGTGCT	753
QY	1536	TCCCGAACACCTTGTGGAATGGTTTACACGACAGCCGCGCAGAGATCGGATTACT	1595
Db	754	TCTTGAGTGTCTTGGANTTGGGCTCACACCAAGCAGCCAGAGAGCTGGGTCTCC	813
QY	1596	CCCCGGCACACCTGTCGCTGTAGGGTTAATTGATGCTCAGCTGTGGCATCGGTACGCT	1655
Db	814	TTACAGGAATTCAGTGGCAGCCTCACTCATTCATCCCATGCGAGGACTAGGATGAT	873
QY	1656	TGGCGTAGAAGTGGAGCGCT-----GAAACAATCT	1685
Db	874	TGGGCAGATGTGAGAGGTCAACGGCTTACCTGTGAGGACAGCCAGCTCAAGGCT	933
QY	1686	CCGCTATGTTTTCGCACTTCTTCATGCACCATGCACTTACCACCTCTCCCTCGTTGT	1745
Db	934	GGCTGTCACTCTGGAACATCATCTTGTACATGGGGATCAACAAAGATCCAGTGTGT	993
QY	1746	ACCGGCTGTCTGGGGCCGATTAACAGTGCATGTTCAGGTCTGTGGTTAGTTGAAGG	1805
Db	994	ACAGGTGTCTGGGACCTTATTTACGCCATGCTCCCTCGGTTCTGTTGAATGAAGG	1053
QY	1806	CGGGCAAAATGCTGGGGAGCAGCTATTGACCACTACTTGTATTTCCATCCGGCTTGA	1865
Db	1054	TGCTCAGAGTGTGACTGGAAATTTGATAGACCATATGTGTGCAGGGCCATCCTGCCTTCC	1113
QY	1866	AGAAGTCGCGAAATGGC	1883

Db	1114	TGACTACAAGCCAAAGC	1131
RESULT 15			
AE009374			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			

11164 bp DNA linear BCT 20-DEC-2001
Agrobacterium tumefaciens str. C58 linear chromosome, section 144
of 187 of the complete sequence.
AE009374 AE008689
AE009374.1 GI:17742928

Agrobacterium tumefaciens str. C58 (U. Washington).
Agrobacterium tumefaciens str. C58 (U. Washington)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 11164)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
and Nester, E.W.
The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
21608550
11743193

2 (bases 1 to 11164)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
and Nester, E.W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA

Location/Qualifiers
1..11164
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
94..1578
/gene="mtlk"
/note="Atu4451"
94..1578
/gene="mtlk"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="mannitol 2-dehydrogenase"
/protein_id="AAL45245.1"
/db_xref="GI:17742929"

/translation="MTCKLSLATLDEAKKTAAPVAYSNDLSAGIVHFGVGNFRAHQ
AVYLDLNTGADHDFATIGAGVLPDSATMRKLAQAQDLFTTVQDNNRGARTVGP
MIDLKVGDTQAIIDTLADPKIRIVSMITTEGGYFIDAGSGFQHPQIAEDGKNPSA
PKTVFGFVAGLKARDDQLQFTVMSCDNIPHNGKVTKNVAVGLAALSDAFANWIG
ENAVFPNSMDRITPATGERNIARDQFDGDNWPFCEEFKQWMDNFPAGRPAL
EKAGVQFVKDVAPEYEHMKIRILNGSHAAIAYPALLDIHFEVHEAMHPLIRAFIAKLE
KDEILPVPVPPDLDNDYFGLIRRELNPKIGTIPRLAODGSGNRQPKFLPSTLDR
LSRGDIVGLSVALMKRCYFYGSDSGKEIVFNDASAEKQAARAKNDPDAFLAL
DDIFGEVASALFRKFRSHALTTLWQEGTAKTLQLYLDLDAEK"

gene

1585..2304

/gene="Atu4452"

CDS

1585..2304

/gene="Atu4452"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45246.1"

/db_xref="GI:17742930"

/translation="MRGTGGLLFDGCVLVDSPEVSIIVLLDMLSHLGVMTGEEEA
YERELGVSAMTTLFEYGVETDIDFLDMRATLFRFETELKPIDGIAETIDELL
PLKRVASSOPPERIRYSLGLIDKFEPHFVFSATVVRNGKPKAPDLFHAARNGVE
PRQIVIEDSPAGIAAKAGMGVFAFTGGSHARFPAPREXIAIGLAGADAVDPMDLV
QLVGSYVGRGGFDQVERDAG"

gene

2359..3936

/gene="Atu4453"

CDS

2359..3936

/gene="Atu4453"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="ribitol kinase"

/protein_id="AAL45247.1"

/db_xref="GI:17742931"

/translation="MRONLVADVGTASARAGIDPAGRIILARSIIHPILMRPRENH
AEPDSTDIWNAVCIAKALADAGVLPQSIAGFDATCSLVIRDERGEPVSVSTDD
DRFDITVWLDHRAIGEADRLTASGHRVLDVFNAGSVSPQMOKMLWKLTHMPQSWSRM
SFATDLADFLTKATGSAOENCTQTAKNFLAQEHGQWQADYLAFAGLDDLKERAGL
PETVMFGATGPIPSPEAAELGLDTGQVAAGMIDAYAGALGALGCLAEADVKGHVA
LIAGTSCLVAMTOPMPSRLGWPYQWALVPHWLVEGGOSATGALLDHIIVRMHAAG
CEPTALHARIVARTLRELEGEAFADRLHLPDPFGNRSPLADPHAVGVVSGLTLD
TSPDLCRLYRTAIALGARHVLDAWREGYAVETLHVGGHVKNPILLMELADVT
KRIIVPATADAVILCTAMTAAATAGGVHASLAAGAAMPNGAEISGNPALAAHYERD
YRRFLAMIRHQELES"

gene

complement(4041..4874)

/gene="Atu4454"

CDS

complement(4041..4874)

/gene="Atu4454"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="transcriptional regulator, RpiR family"

/protein_id="AAL45248.1"

/db_xref="GI:17742932"

/translation="MVEKLLNLTRAMKAGTPSPERKIAKYLIEHLDELFFETAOTLAA
KLUSPMTVGRFLSLGYRQFSDIRADLRHAETVVGADQPVTAEGGRHSNPFSQLLS
QOIQAIQTAFDMTQTQIWRLAMNETTSADDVCLATSPGLDAGRHFYGRLEECRHHI
YLGSDNAYVALWDYDIPARTLLVIMDCGNNLLPLQRLSTTARKGKYRTLITTFYEW
GPESVDLCLAMPQSONGGQLQLVSLLEFTLICALSAGADDAGRKARVKNLAALKRSLR
S"

gene

4897..5067

/gene="Atu4455"

CDS

4897..5067

/gene="Atu4455"

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAL45249.1"

/db_xref="GI:17742933"

/translation="MNGDAFAPCKRANHDLTVETSPFVHLHRTGMRIIVDGLLYFR
AIFYSRIPEFAV"

gene

5170..5907

/gene="Atu4456"

CDS

5170..5907

/gene="Atu4456"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45250.1"

/db_xref="GI:17742934"

/translation="MIFDAARLAFENLFAAETRSVFWKVLGTLVLVLAALFAIRSF
IYFALPWIDTLIPGVPDWAGLSFVFAIFAGIGLALALLISPTAVTAGLFLDDVA
DVVEKQNYPPDDPGKAMPDIEAVLSSRIAGNLVALLLLVPGVNLIAFFLVN
GYLLGREFFEFAAMRFPADARLFRVTRKFTVFMAGLVIAFLAVPFLNLITPLFAA
SLMVHLHKAVSRDPSFAAGGTEQLRG"

gene

complement(5869..6579)

/gene="Atu4457"

CDS

complement(5869..6579)

/gene="Atu4457"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="endonuclease III"

/protein_id="AAL45251.1"

/db_xref="GI:17742935"

/translation="MRVKTAYSKDELTEIFRRFSIORPEPKGELEHTNPFTLLVAVAL
SAOATDGVNRAIRALFKVADTPKMLALGEOLIGHIKTIGLYRNKAKNVLTALSOML
IDNFGGEVPKTRREELVTLPGVGRKTAIVYMSAFQVPTLAVDTHVFRANRLCLAPGK
TTDEVEDRLRIIPQYLFAHHLWILHGRYCCARKPECCRCVADIADICKSPEKTFDI
PAPLVELPAQLFGAAGG"

gene

6689..7171

/gene="Atu4458"

CDS

6689..7171

/gene="Atu4458"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45252.1"

/db_xref="GI:17742936"

/translation="MDTLNDQPIFAELVPHRSRGRFRLLIITGTACLIYGGFFL
ATGAWPILGFLGDLFLLYIAFRANRAAKAREEVSRTNLSIRKSPAGRMVYRE
NPFWARFVRHDEIGIVSMHVTGGRATDIGSFLNPDDRNSFAKAFGALATVRRRM
"

gene

7438..8310

/gene="ada"

CDS

7438..8310

/gene="ada"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="6-O-methylguanine-DNA methyltransferase"

/protein_id="AAL45253.1"

/db_xref="GI:17742937"

/translation="MNANIMLNEITPIGSDYDVTVRGVIETLLTDYRQPSLEATAAR
LGQSPQLOKQFTFRWAGLSKRAFLQAVTLDRHAKRLRLREEDLPLETSEVHSGFGRLL
HDLFVTHEAMSPGKWKAGGGLTIYGFHASPGLALVMTDRGLAGCAFAFDGDERA
CFEDMAGRPNADYVEDREATAFYAARIFEPAMWTADKPLRVLLGTDFQVFWKSL
KIPMGRAVTVNIACDIIQGTASRAVGAAGVNPVSFVVPCHRAVKGSGALTGYHWGL
TKRAMLGWETGA"

gene

8375..8599

Query Match	5.5%	Score 218.6;	DB 1;	Length 11164;
Best Local Similarity	49.3%	Pred. No. 5.2e-52;		
Matches 714;	Conservative 0;	Mismatches 694;	Indels 39;	Gaps 4;
QY	872	AAACCGTTATTGCTGTGATGTCGGATCAGGAGTCTCCGCCGCCGGGATTTTGA	CTCA	931
Db	2369	AAAAATCTCGTCTGCTGCGATGTCGGTACCGGAGCGCGCGCGGATTTTCGATCC	AG	2428
QY	932	ACGGATCTCTGCTATCCCATGCCAGAAAAATACGACTACGCGCGCGAGCGGAGCC		991
Db	2429	CGGGCGGCTGTCGGCGGCTCCATTCATCCCATCTCATGCGAGCGCGCGGAAACC		2488
QY	992	GCGTGAACAGTCCAGCCAGGAGATCTGCAGCGGCTGCTGTTTCATGATTCGTAATGC	GC	1051

```
Db 2489 ACCCGAGCAGATCCACGATATATGGAAATGCCGTGTCATTTGCCGTGAAGCGCGCGC 2548
QY 1052 TCACCTCGGCAGACGTTTGTGCAAAAGTGTGGCAGCATCGGTTTGTATGCCACCTGTT 1111
Db 2549 TTCCCGATCGCGCGTCTGCTCCGCAAGCATTCGCCCATTTGTTTGACGCCACCTGCT 2608
QY 1112 CTCTGTGTGTGATGATAAAACCGGTGATGCCATTGCTGTACGCCGGAAGAGATGCAA 1171
Db 2609 CGCTGTGTATCCCGATGAGCGGGGAGCCGTTTTCGTTTTCACGACCGATGACGACC 2668
QY 1172 AGCAAAATATCATGTCTGGATGATCACCGCGCCACCGCAACAGCAGAGCAATCAATG 1231
Db 2669 GTTTCAGACCATCTGCTGTGGCTTGACACCGCGCCATCGCGAGCGCGACCGGCTGACGG 2728
QY 1232 CCACCTCACCATCGGTGCTGAACTAGCTGGTGGTAAATTTCCGCTGAAATGGAACAC 1291
Db 2729 CCTCGGGCCATCGGTGCTCGATTTCCCGGCAACAGCGCTCTCGCCGGAATGCAGATGC 2788
QY 1292 CGAAATTTCTCGGCTGAAGAAATATGCCAGAGATCTACGAACGTGCCGGACAATTTT 1351
Db 2789 CGAAGCTGATGTGGCTGAAGACGCATATGCCGCAAAAGCTGGTCGCGCATGCTCCCTGC 2848
QY 1352 TCGATCTGGCCGATTTCTGACCTGCGGCTACCGGTGATTTAGCGCTTCAGTATGCA 1411
Db 2849 TCGATCTCGCGGATTTCTGACCTGGAAGCGACCGGCTCGGCAACAGCGCTCCCAATTGCA 2908
QY 1412 CTGTTACCTGTAAATGGACGTGGCTGGCACATGAAATGCTGGGATCCAGATTAATTTCC 1471
Db 2909 CGCAGACGGGAAATGGAATCTCTGGCGAGCAACATCCCGCTGGCAGCGCGATATC 2968
QY 1472 GCACCATTTGCGCTTGCGAGATTAGCGGATGAAGATTTTATTCGCATTTGTCATCATTTG 1531
Db 2969 TGGCTTTTGGCGACTGGAGACCTGAAGAGCGGCGAGGCTGCTGAAACTACCGTCA 3028
QY 1532 TTTTCTCCGGAACACCTTGTGGAATGGTTTAACACACAAGCCGCGGAGAGATGGAT 1591
Db 3029 TGGCGGGCGGCGCATCGG-----CCGCTTTTGGCCGAGGCGCGCGAGAACTTGGTC 3082
QY 1592 TACTCCCGGCACACCTGTGCTGTAGGGTTAATGTATGCTACGCTGTGGTGGCATCGGTA 1651
Db 3083 TCGATACCGGCTCCAGGTGGCGCGGCGCATGATGCTATGCCGTGCGCTCGGG 3142
QY 1652 CGGTTG-----CGGTAGAGGTGGAGCGCTGAACAATCTCGGATATGTTTTCGGCACAT 1705
Db 3143 CTCTGGGTGGATGCTTGGCGAGGATGTGCGTAAACATGTGCGGCTCATCGCGGCGACGT 3202
QY 1706 CTTTCATGCACCATGGCATCTACCACCTTCTCCCTGTTTGTACCGGCTGTCTGGGGCCGT 1765
Db 3203 CGAGTTGCTTGTGCGCATGTCCACGACCGCATGCCGCGCGCAGCTGTGGGGACCCCT 3262
QY 1766 ATTACAGTGGATGGTTCCAGGTCTGTGGTTAGTTGAAGCGGGCAAAAGTGTCTCGGGGAG 1825
Db 3263 ACTGGCAGGGGGTGTGCTTGGCCCTGGCCACTGGCTGTGGAGGGTGGGCAATCGGCTACCGGG 3322
QY 1826 CAGCTATTGACCAAGCTACTTGTATTTCCATCCGCGCTGTGAAGAAAGCTCGCGAAATGGCAC 1885
Db 3323 CGCTGTCTGACCATATCGTGGCATGCGGCGGATGCGGCGGGTGGCGA-----GCCGG 3370
QY 1886 AACGTGTGAATCAGCCCCCTCCCGCTTGGCTTGTCTGATCGAATCTCGAAAAACGGCGC 1945
Db 3371 ATACGGGCTTGCATGCCCGCATGTTTGGCGCGCTCACTGAATTTGCCGGAACCTGGAGGGG- 3429
QY 1946 AACCATCAGATGCTGTGGCCCTGGCGAAAGGGCTACACGTGGTGGCGGAATTTCTCGGAA 2005
Db 3430 -----GAGGCTTTGGCGGACCGGCTGATGTTCTGGCTGACTTCCACGGCA 3475
QY 2006 ATCGCGCCCGCTTCGCAGATCTCATGCGCAGACGGTAATTTGTGGCCCTGGGTATGGAGC 2065
Db 3476 ACCGTTTCGCGCTTGGCGGATCCGCATCGCGTCTGGCGTCTGAGCGGCTGACGCTCGATA 3535
QY 2066 GAGATCTGGATAATTTACTCGCCTTCTATATCGCTGGATTATCGCGGAATTTGTTATGGTC 2125
```

```
Db 3536 CTTCTCTCGACAGTCTCTCTGCCGTCTTTATTTGGCGCACCGCGATCGCGATCGCGCTCGGCG 3595
QY 2126 TGGCCAAAATTTCTGACGCTCAAAACAGCGCAGGAGTAGTGAAGTAAATAATATCGTTATTA 2185
Db 3596 CGCGCCACGCTGCTCGATGCCATGGAGCGTTTCGGTTATGCGGTGGAACCGCTGCATGTTA 3655
QY 2186 GCGGCGGTGCGGCGAGCATCCACTGTGTACGACAAATTTCTGGCAGATACCTGCGGTATTC 2245
Db 3656 CCGCGGGCATGTGAAGAACCGCTGTGTATGAGGCTTTACGCCGATGTACGCGCAAGC 3715
QY 2246 CGGTCAATTACCACGAATGCTGCGAAACCGGTTTTATTAGGCTCGGCATTTCTTGGTCTG 2305
Db 3716 GCATCGTGGTTCCGCGACCGCGGATGCGGTGCTGCTCGGCACGGCATACGCGCGCA 3775
QY 2306 TCGCCGG 2312
Db 3776 CGGCGG 3782
```

Search completed: March 21, 2003, 20:42:22
Job time : 11847.3 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	3971	100.0	3971	22	AA016811	<i>Escherichia coli</i> s
2	311	7.8	1401	24	ABN90720	<i>Staphylococcus epi</i>
3	302.4	7.6	1876	24	AA038854	Human kinase (PKIN
4	302.4	7.6	2050	22	AA0159449	Human polynucleoti
5	300.8	7.6	1991	22	AA0303016	Human diagnostic a
6	211.6	5.3	2602	23	AA026431	<i>Drosophila melanog</i>
7	183	4.6	1947	22	AAH15580	Human cDNA sequenc
8	177.4	4.5	5019	23	AA026430	<i>Drosophila melanog</i>
9	124	3.1	1044	22	AAH14791	Human cDNA sequenc

XX
PF
08-MAR-2001: 2001WO-US07A74

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted and is derived by analysis of the total score distribution.

Db 1561 TTAACAGCACAAAGCGCGCAGAGATGGGATTACCTCCCGGCGACACCTGTCTGCTGTAGGG 1620
QY 1621 TTAATTTGATGCTACAGCTGCTGGTGGCATCGGTACGGTTGGCGTAGAAGGTGGAGCGCTGAAC 1680
Db 1621 TTAATTTGATGCTACAGCTGCTGGTGGCATCGGTACGGTTGGCGTAGAAGGTGGAGCGCTGAAC 1680
QY 1681 AATCTCGGCTATGTTTTCGGCACTTCTTCATGCGACATGCGCATCTACCACTTCTCCCTCG 1740
Db 1681 AATCTCGGCTATGTTTTCGGCACTTCTTCATGCGACATGCGCATCTACCACTTCTCCCTCG 1740
QY 1741 TTTTGTACCGGGTGTCTGGGGCGGTATTACAGTGCAGATGGTTCCAGGTCTGTGGTTAGTT 1800
Db 1741 TTTTGTACCGGGTGTCTGGGGCGGTATTACAGTGCAGATGGTTCCAGGTCTGTGGTTAGTT 1800
QY 1801 GAAGCGGGCAAAAGTCTCGGGGAGCAGCTATTGACCAAGCTACTTGATTTCCATCCGGCT 1860
Db 1801 GAAGCGGGCAAAAGTCTCGGGGAGCAGCTATTGACCAAGCTACTTGATTTCCATCCGGCT 1860
QY 1861 GTTGAAGAAGCTCGGAAATGCGCAACAGTGTGAATCAGCCCTCCCGTCTGGCTTGTCT 1920
Db 1861 GTTGAAGAAGCTCGGAAATGCGCAACAGTGTGAATCAGCCCTCCCGTCTGGCTTGTCT 1920
QY 1921 GATCGAATCTCGAAAAACGGCGCAACCATCAGATGCTGTGCGCCCTGCGCAAAAGGGCTA 1980
Db 1921 GATCGAATCTCGAAAAACGGCGCAACCATCAGATGCTGTGCGCCCTGCGCAAAAGGGCTA 1980
QY 1981 CACGTGGTSCCGGAATTTCTCGGAAATCGCGCCCTTCGAGATCTCTCATGCCAGAGCG 2040
Db 1981 CACGTGGTSCCGGAATTTCTCGGAAATCGCGCCCTTCGAGATCTCTCATGCCAGAGCG 2040
QY 2041 GTAATTTGTGGCTGGGTATGAGGAGGAGATCTGGATAATTTACTGCGCTTGTATATCGCT 2100
Db 2041 GTAATTTGTGGCTGGGTATGAGGAGGAGATCTGGATAATTTACTGCGCTTGTATATCGCT 2100
QY 2101 GGATTTATGCGGAATTTGTTGTTGCTGTGCGCAAAATTTCTGACGCTCAAAACAGCGAGGA 2160
Db 2101 GGATTTATGCGGAATTTGTTGTTGCTGTGCGCAAAATTTCTGACGCTCAAAACAGCGAGGA 2160
QY 2161 GTAGTAGTAAATAATATCGTTATTAGCGCGGTGCGGCGAGCATCCACTGGTAGACAAA 2220
Db 2161 GTAGTAGTAAATAATATCGTTATTAGCGCGGTGCGGCGAGCATCCACTGGTAGACAAA 2220
QY 2221 ATTCTGGCAGATACCTGGCGGTATTCCGGTCAATACCAGCAATGCTGCGAAACGGTTTFA 2280
Db 2221 ATTCTGGCAGATACCTGGCGGTATTCCGGTCAATACCAGCAATGCTGCGAAACGGTTTFA 2280
QY 2281 TTAGGCTCGGCCATTCTTGGTGTCTGCGCGGAATAATTTGACACTTCTCTTGTGGCGAAGCG 2340
Db 2281 TTAGGCTCGGCCATTCTTGGTGTCTGCGCGGAATAATTTGACACTTCTCTTGTGGCGAAGCG 2340
QY 2341 ATGCAACAATTCACCCATGTGGATAAATAATTATTATCCGCAAGACGCTATCAATCTCTT 2400
Db 2341 ATGCAACAATTCACCCATGTGGATAAATAATTATTATCCGCAAGACGCTATCAATCTCTT 2400
QY 2401 CATCATGTCGATATGAGCCCTATAGCAGTTGCGAGCATACTGCAAAATTAATCTCAGAGAC 2460
Db 2401 CATCATGTCGATATGAGCCCTATAGCAGTTGCGAGCATACTGCAAAATTAATCTCAGAGAC 2460
QY 2461 TAAATAACAGCGCTGACGCTGTTTTCAGGCAATCACTAATACGACTCACTCCGGTA 2520
Db 2461 TAAATAACAGCGCTGACGCTGTTTTCAGGCAATCACTAATACGACTCACTCCGGTA 2520
QY 2521 ATATCCGGAGTGCATTTCACTGCAACCTTAAACAGAGTTTATATGTCGCAAAATAATA 2580
Db 2521 ATATCCGGAGTGCATTTCACTGCAACCTTAAACAGAGTTTATATGTCGCAAAATAATA 2580
QY 2581 AACAGTGGTTGGTTGGCACTGCACTGTGATATGGGATATATCGCCATCGCAGTATTTA 2640
Db 2581 AACAGTGGTTGGTTGGCACTGCACTGTGATATGGGATATATCGCCATCGCAGTATTTA 2640
QY 2641 TCAGTGGTATGGTTTCGAACTCGCAATCTTATCGCACTATATTAATCGTTAGGCTTCA 2700
Db 2641 TCAGTGGTATGGTTTCGAACTCGCAATCTTATCGCACTATATTAATCGTTAGGCTTCA 2700

QY 2701 CACCTGCGGAAGCCTCTTTTGCTTTACGCTCTACGCGCTGGCGCTGCCCTTTTCGCGCT 2760
Db 2701 CACCTGCGGAAGCCTCTTTTGCTTTACGCTCTACGCGCTGGCGCTGCCCTTTTCGCGCT 2760
QY 2761 GGGTTTCTGGGTAGTAGGGAATATCAGCCCGCAAAACCAATGCTGATGGTTTGG 2820
Db 2761 GGGTTTCTGGGTAGTAGGGAATATCAGCCCGCAAAACCAATGCTGATGGTTTGG 2820
QY 2821 TCCTATGGTGGCTATTCCCATGTTCTGTTCTGGTCTTTGGATTAGGACAGGCAAACTATG 2880
Db 2821 TCCTATGGTGGCTATTCCCATGTTCTGTTCTGGTCTTTGGATTAGGACAGGCAAACTATG 2880
QY 2881 GGTAAATCTGCTGTTTACGGGATTCGGTTCAGGTATCGGTATCTCTACTCAT 2940
Db 2881 GGTAAATCTGCTGTTTACGGGATTCGGTTCAGGTATCGGTATCTCTACTCAT 2940
QY 2941 TTATCGTCTGTTTATTTATCATACGTCGTAGCGAAATTTCCAGTTCTGCTCTGGTGGT 3000
Db 2941 TTATCGTCTGTTTATTTATCATACGTCGTAGCGAAATTTCCAGTTCTGCTCTGGTGGT 3000
QY 3001 ACTGGCGGTATATTTCAGTAGGATCGGTGTTGCTGGCAGTTATATTTCCAGTTTACGA 3060
Db 3001 ACTGGCGGTATATTTCAGTAGGATCGGTGTTGCTGGCAGTTATATTTCCAGTTTACGA 3060
QY 3061 TACCGATTATGGGTGAATGGGAACCTTATGGCTGGCAGTGGCGTTCGTGTTGCTGGCG 3120
Db 3061 TACCGATTATGGGTGAATGGGAACCTTATGGCTGGCAGTGGCGTTCGTGTTGCTGGCG 3120
QY 3121 GTGTCAATGCCATGATCTCTTCCGTCATGTTAAACGCCCTGGACATATGCAATAATTTAA 3180
Db 3121 GTGTCAATGCCATGATCTCTTCCGTCATGTTAAACGCCCTGGACATATGCAATAATTTAA 3180
QY 3181 CTCGCCGTGAGAAATTTGAGAAATTAAGTCGGGCACTAACTTTACTTTATACCAACGCA 3240
Db 3181 CTCGCCGTGAGAAATTTGAGAAATTAAGTCGGGCACTAACTTTACTTTATACCAACGCA 3240
QY 3241 ATATTTTCTCTCCAGTATTTGCGCATATATAAATACCTTATCTGTTATTCGGTTTGGCG 3300
Db 3241 ATATTTTCTCTCCAGTATTTGCGCATATATAAATACCTTATCTGTTATTCGGTTTGGCG 3300
QY 3301 TCATTTATGCCAATGATTTTGGATGAACCTGGATTCACCACTCTGAATGCTGCGAG 3360
Db 3301 TCATTTATGCCAATGATTTTGGATGAACCTGGATTCACCACTCTGAATGCTGCGAG 3360
QY 3361 TCTGGCGGCATTTTCTTTACCACTATTTTCTCTAATATTTTGGGGAATGTCGAG 3420
Db 3361 TCTGGCGGCATTTTCTTTACCACTATTTTCTCTAATATTTTGGGGAATGTCGAG 3420
QY 3421 AAAAAATGGCTGGATGCTGTTTATTCGCTGGTTCGCTCGGATGGCGATGGCAGATCAA 3480
Db 3421 AAAAAATGGCTGGATGCTGTTTATTCGCTGGTTCGCTCGGATGGCGATGGCAGATCAA 3480
QY 3481 GTTTAGCTGTTTACTACATGCGCAATCTTTGGTTCGCACTACTGGATGGCAATGATTC 3540
Db 3481 GTTTAGCTGTTTACTACATGCGCAATCTTTGGTTCGCACTACTGGATGGCAATGATTC 3540
QY 3541 CGCGGATTCCTGCGAATTTTGTGCTGCAATTTGTCGCGATGGCGCTGCTTCCCGG 3600
Db 3541 CGCGGATTCCTGCGAATTTTGTGCTGCAATTTGTCGCGATGGCGCTGCTTCCCGG 3600
QY 3601 CACTGGAACCAAAACAAAGGTGCTGCAATCTCGGTTTACAACTCTCTCGCGGTATGT 3660
Db 3601 CACTGGAACCAAAACAAAGGTGCTGCAATCTCGGTTTACAACTCTCTCGCGGTATGT 3660
QY 3661 CTAACTTCCTGGCTCGCGCAATGGCGTGGTGTATTACCGTGGTTTAGCACTATCGGTG 3720
Db 3661 CTAACTTCCTGGCTCGCGCAATGGCGTGGTGTATTACCGTGGTTTAGCACTATCGGTG 3720
QY 3721 TGGTCAATTCCTATACAGCATTTGATCTATTTGGCTTTTCTGCTTTGGCAATTCATTCGCG 3780
Db 3721 TGGTCAATTCCTATACAGCATTTGATCTATTTGGCTTTTCTGCTTTGGCAATTCATTCGCG 3780

The invention relates human kinases (PKIN) and their corresponding nucleic acid sequences. PKIN and its DNA are useful for diagnosing, treating and preventing cancer, an immune system disorder (e.g., acquired immune deficiency syndrome (AIDS), Addison's disease, allergy asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders

Db 508 TTTTTTGGATAGCAGTTTACCCCATTTACCATCAACAGGAAGGGATTCCTCCATCGAAA 567
Qy 1179 TATCATTTGTGGATGATACCGCGCCACCAAGCAAGCAGCAATCAATGCGCACATCA 1238
Db 568 CGTCATCATGTGGTGACCATCGAGCAGTCAGTCAAGTTACAGGATCAATGAGACCAA 627
Qy 1239 CCATCCGGTGTGAACACTACGTGGTGGTAAATTTTCGCTGAAATGGAACACCGCAAAAT 1298
Db 628 GCACAGTGTCTCCAGTACGTGGGGGGGTGATGCTGTGGAATGCGAGGCCCGCAAACT 687
Qy 1299 TCTCTGGCTGAAGAAATATGCCAGAGAT---CTACCAACGTGCCGGACAATTTTTCGA 1355
Db 688 TCTGTGGCTGAAGAGAACTTTGAGAGAGATTTGCTGGGATAGGCGGGACATTTCTTTTGA 747
Qy 1356 TCTGGCCGATTTTCTGACCTGGCGGCTACCGGTGATTTAGCGCTTCAGTATGACATGT 1415
Db 748 TCTCCAGACTTCTATCGTGAAGCAACAGGTGTACAGCAGCGTCTCTGTCGCCCT 807
Qy 1416 TACCTGTAATGGAGCTGGCTGGCACATGAAATCGCTGGATCCAGATTTATTTCCGCAC 1475
Db 808 GGTGTGAAGTGACATATTACGCA---GAGAAAGCTGGGACGACAGTTTCTGGAATAAT 864
Qy 1476 CATTTGGCTTGCAGAGTTAGCGGATGAAGATTTTATTCGCATTTGTCATCATATTTTTC 1535
Db 865 GATTTGGTTTGAAGACTTTTGTTCAGATAATTACAGCAAAATAGGAAACCAAGTGTACT 924
Qy 1536 TCCCGGAACACTTCTGGAATGFTTAAACAGCAACGCCGCGCAGAGATGGGATTACT 1595
Db 925 TCTTGAGCTTCTCTTGGAAATGGCTCACACGAGGACAGACAGTGTGGCTTCT 984
Qy 1596 CCCCCGACACTGTGCTGTAGGGTTAATTGATCTCACGCTGTGTGCTGCTGCTGCTGCTGCT 1655
Db 985 CCTGGGATTCGGTGCAGCTTCTACTTATGATGCCCATGCAGGAGGACTAGGAGTGTAT 1044
Qy 1656 TGGCTAGAAGTGGAGCGC-----TGAAACAATCT 1685
Db 1045 TGGGCGAGATGTAGAGGGGACGGCCCTCATCTGTGAGGGGACGCCAAGTGCACGCT 1104
Qy 1686 CCGGTATGTTTTCGGCACTTCTTACCATGATGATGATGATGATGATGATGATGATGATGAT 1745
Db 1105 GGCTGTCATCTGTGAAGCTTCTTGTGCATGGGGATCAGCAAGACCCGATTTTGT 1164
Qy 1746 ACCGGGTGTCTGGGGGGCGTATTAACAGTGCAGTGTCCAGGTCTGTGTAGTTGAAGG 1805
Db 1165 ACCAGCGCTCTGGGGCGCTTATTTCTCAGCCATGTTACCTGGGTCTGCTGAATGAAGG 1224
Qy 1806 CGGGCAAGTGTCTGGGAGCAGCTATTGACCAAGCTACTTGTATTCATCCGCTGTTGA 1865
Db 1225 TGGTCAGAGCGTTACTGAAATTTGATAGACCACATGATGACAGGCCATGCTGCTTTTCC 1284
Qy 1866 AGAAGCTCGGAAATGGCACAACGTGTGAATCAGCCCTCCCGCTGCTGCTGCTGCTGCTG 1925
Db 1285 AGAAGCTAAGTAAGGCCACAGCCAGATGCCAGATATATGAT---ATTGAAACAG 1341
Qy 1926 AATCCTCGAAACCGGCGCAACCATCAGATGCTGCTGCCCTGGCGAAAGGCTACAGCT 1985
Db 1342 TCACCTGATCTGATTAAGAAGGCTCAGCCTGTGGGTTTCCCTACTGTTGATTTACATGT 1401
Qy 1986 GGTCCGGAATTTCTCGAAATCGCGCCCTTCCAGATCCTCATGCCAGAGCGGTAAAT 2045
Db 1402 TTGGCCGATTTTCCATGCAACCGGCTCTCCCTTAGCAGATCTGACACTTAAAGGGCATGT 1461
Qy 2046 TTGTGCTGGGTATGGAGCGAGATCTGGATAATTTTACTTCGCTTGTATATCGCTGGATT 2105
Db 1462 CACCGGATTAAGTGTCTCAGGACCTTGATGATCTTGCCATTTCTTCTACCTGGCCACAGT 1521
Qy 2106 ATCGGGAATTTGTTATGCTCTCGGCAAAATTCGACGCTCAACAGCGGAGGAGTGTAGT 2165
Db 1522 TCAAGCCATTTGCTTTGGGAGCTCGCTTCTATTATAGAAGCCATGGAGGACGAGGCACTC 1581
Qy 2166 GAGTAAAAATATCGTTTATACGGGGGTCGCGGCGGATCATTCAGTGTACGCAAAATCT 2225

Db 1582 AATCAGTACTCTTTTCCCTATGTGGAGGCTCAGCAAGAAATCCCTTTTGTGCAAAATCA 1641
Qy 2226 GCGAGATACCTGCGGTATTTCCGGTCAATACCAACCAATGCTCGCAACCGGTTTATTAGG 2285
Db 1642 TCGGACATTAATGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1701
Qy 2286 CTCGGCCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2345
Db 1702 TCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1761
Qy 2346 ACAATTCACCATGCGGATAAAT 2405
Db 1762 AAAATGAGCAAGTTGGAAAGTTGTGTTCCCGAGACTACAGGATAAAAAATACTATGA 1821
Qy 2406 TCGTGCATATAGGCTATATAAGCAGTTG 2433
Db 1822 TAAGAATACCAAGTATTCCTGAAGCTG 1849
RESULT 5
ID AAS03016 standard; cDNA; 1991 BP.
XX AAS03016;
AC AAS03016;
DT 29-AUG-2001 (first entry)
XX Human diagnostic and therapeutic (dithp) cDNA sequence #5.
XX Human diagnostic and therapeutic molecule; dithp; gene therapy;
KW thalassemia; cardiovascular disorder; cell proliferative disorder;
KW cancer; neurodegenerative disorder; autoimmune disorder; enzyme;
KW infectious disorder; inflammatory disorder; developmental disorder;
KW Incyte ID number 05377811dec; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO200121836-A2.
XX 29-MAR-2001.
XX 19-SEP-2000; 2000WO-US25643.
XX 23-SEP-1999; 99US-0155760.
PR 24-SEP-1999; 99US-0155939.
PR 24-SEP-1999; 99US-0156294.
PR 28-SEP-1999; 99US-0156565.
PR 28-SEP-1999; 99US-0156624.
PR 28-SEP-1999; 99US-0156625.
PR 24-NOV-1999; 99US-0167410.
PR 24-NOV-1999; 99US-0167453.
PR 24-NOV-1999; 99US-0167517.
PR 24-NOV-1999; 99US-0167520.
PR 24-NOV-1999; 99US-0167542.
PR 29-NOV-1999; 99US-0167943.
PR 29-NOV-1999; 99US-0167945.
PR 30-NOV-1999; 99US-0168197.
PR 30-NOV-1999; 99US-0168265.
PR 30-NOV-1999; 99US-0168429.
PR 30-NOV-1999; 99US-0168432.
PR 01-DEC-1999; 99US-0168468.
PR 01-DEC-1999; 99US-0168599.
XX (INCY-) INCYTE GENOMICS INC.
XX Hodgson DM, Lincoln SE, Russo PD, Spiro PA, Banville SC;
PI Bratcher SR, Dufour GE, Cohen HU, Rosen BH, Shah P, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR;
PI Roseberry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK;
PI Amshey S, Fong WT;
XX WPI; 2001-281607/29.
XX

PT Novel diagnostic and therapeutic polynucleotides, used in disease
PT diagnosis and for gene therapy of conditions such as cancer and
PT thalassemia

PS Claim 1; Page 253-254; 299pp; English.

XX The present sequence for human diagnostic and therapeutic (dithp) cDNA
CC sequence #5 is 1 of 71 (AA503012-AA503082) novel sequences described
CC in the invention. The present sequence (Incye ID No: 05377811dec)
CC encodes an enzyme molecule. The dithp polynucleotides may be used
CC to diagnose a condition disease or disorder associated with human
CC molecules. They can be used to identify the presence of similar
CC nucleic acids. Dithp polynucleotides may be used to generate hybridisation
CC probes for use in chromosomal mapping. Polypeptides (dithp) encoded by
CC dithp are used to screen for molecules which bind to them and modulate
CC their activity. Dithp polynucleotides can be used for gene therapy of
CC disorders such as severe combined immunodeficiency syndrome (SCID),
CC cystic fibrosis, thalassemia, haemophilia resulting from Factor VIII
CC or IX deficiencies, cardiovascular disorders, e.g familial
CC hypercholesterolaemia (FH), cell proliferative disorders e.g. cancers,
CC neurodegenerative disorders, autoimmune/inflammatory disorders,
CC infectious disorders and developmental disorders. The antibodies can be
CC used to analyse protein expression levels.

XX Sequence 1991 BP; 468 A; 467 C; 534 G; 516 T; 6 other;

Query Match 7.6%; Score 300.8; DB 22; Length 1991;
Best Local Similarity 52.0%; Pred. No. 1.3e-84;
Matches 826; Conservative 0; Mismatches 717; Indels 45; Gaps 5;

QY 879 TATTGTGTGTGATGGGATCAGGAGTGTCCGCGCGGGATTTTGTATCTCAACGGATC 938
DB 227 TGTGGGTGTGGAGTTGGAAACAGGCACTGTCCGTGACGCTTGGTGACAGAGTGGGT 286
QY 939 TCTGCTATCCCATGCCACAGAAAAATACGACTACGCGCGGCAGCGGAGCGCGTGA 998
DB 287 CCTGTTGGCTTTTGCAGACCGCAATTAAGAAATGGGAGCCCGCAGTTCAACACCATGA 346
QY 999 ACAGTCCAGCCAGGAGATCTGCGAGCGGTCTGTTTCATGTATCGTAATCGCGTCACTCT 1058
DB 347 GCGATCCTCCGAGGACATCTGGCTGCGTGTGTGTGTCACAAAGAAAGT-----TGT 400
QY 1059 GCGAGACGTTTGTGCACAAAGTGTGCGCAGGATCGGTTTGTATGCCACCTGTCTCTGTT 1118
DB 401 ACAAGGGATTTGATTTAAACCAAAATTCGAGGACTTGGGTTTGTATGCCACGCTTCTCTGT 460
QY 1119 GGTACTGGATAAAACGGTGATCCATTGCTGTACGCCCGGAAGGAGATGCAAGCAAAA 1178
DB 461 TGTGTTGGATNAGCAGTTTACCCATACAGTCAACCAAGTAAGGGGATTCGCCATCGAAA 520
QY 1179 TATCATTTGTGTGGATGGATCACCGCCCGACCGAAGCAGAGCGCAATCAATGCCACTCA 1238
DB 521 CGTCATCATGTGCTGGACCATCGAGCATCAGTCAGTCAAGTTAACAGGATCAATGAGACCAA 580
QY 1239 CCATCCGGTGTGAACTAGCTGGTGGTAAATTTCCGCTGAAATGGGAACACCGGAAAT 1298
DB 581 GCACAGTGTCTCCAGTACGTGCGGGGGGTGATGTGTGTGAAATGCAGGCCCGCGAAACT 640
QY 1299 TCTCTGGCTGAAGAAATATGCCAGAT---CTACGAACGTGCCGGACAAATTTTTCGA 1355
DB 641 TCTGGCTGAAGAGAACTTGAGAGAGATTTGCTGGGATAGCGGGGACATTTCTTTGA 700
QY 1356 TCTGGCCGATTTTCTGACCTGGCGGCTACCGGTGATTTAGCGGCTTCAGTATGCACTGT 1415
DB 701 TCTCCGGGACTTCTTATCTGTTGAAAGCAACAGGTGTCAGCAGCGGTCTCTGCTCCCT 760
QY 1416 TACCTGTAATGGACGTGCTGCACATGAAATCCCTCGGATCCAGATTTATTCGGCAC 1475
DB 761 GGTGTGTAAGTGSACATATTCAGCA---GAGAAAGGCTGGGACGACAGTTTCTTGGAAAA 817
QY 1476 CATTGGCTTGCAGAGTTAGCGGATGAAGATTTTATTCGATTTGGTCATCATATTTCTTC 1535
DB 818 GATGGTTTGGAGACTTTGTTGCAGATAATTACAGCAAAATAGGAAACCAAGTGTCTACC 877

QY 1536 TCCCGGAACACCTTGTGAAATGGTTTAAACAGCACAAAGCCGCGCAGAGATGGGATTACT 1595
DB 878 TCCTGGAGCTTCTCTTGGAAATGGGCTCACACAGAGGAGCAAGAGACTTGGCCTCT 937
QY 1596 CCGCGCACACCTGTGCTGTAGGGTTAAATTGATGTCTAGCTGGTGGATCCGTTACGGT 1655
DB 938 CCTCGGATTTGGGTCGCGAGCTTCACTCATATGATGCCCATGCAGGAGGACTTGGAGTAT 997
QY 1656 TGGCGTAGAAGTGGAGCGC-----TGAAACAATCT 1685
DB 998 TGGGCGAGATGTGAGAGGCGACGGCCCTCATCTGTGAGGGCAGCCAGTACGTCACGGCT 1057
QY 1686 CCGCTATGTTTTCGGCACTTCTTCATGCACCATGGCATCTACCACTTCTCCCTCGTTGT 1745
DB 1058 GCTGTCTATCTGTGGAACCTCTTCTTGTACATGGGATCAGCAAAAGCCGATTTTGT 1117
QY 1746 ACCGGTGTCTGGGGCCCTATTTACAGTCCGATGGTTCCAGGCTCTGTGTAGTCTCAAGG 1805
DB 1118 ACCAGCGCTCTGGGGCCCTTATTTCTACGCCATGGTACCTGGGTTCTGTGAATGAAGG 1177
QY 1806 CGGGCAAACTGTGCGGGAGCAGCTATTGACCAAGCTACTTGTATTCATCCGCGCTGTGA 1865
DB 1178 TGTGACAGCGTTACTTGGAAATTTGATAGCCACATGTTACAGGCCATGCTCTTTCC 1237
QY 1866 AGAAGCTCCGAAATGGCAACAGTGTGAATCAGCCCCCTCCCGCTCTGGCTGTCTGATCG 1925
DB 1238 AGAAGTACAAGTAAAGGCCACAGCCAGATGCCAGAGTATATATGCAAT---ATTGAACAG 1294
QY 1926 AATCTCTGAAAAAAGCCGCAACCATCAGATGCTGCGCCCTGGCGAAGGGCTACAGCT 1985
DB 1295 TCACCTGGATCTGATTAAGAAGGCTCAGGCTGTGGGTTTCTTACTGTTGATTATCATGT 1354
QY 1986 GGTGCGGAAATTTCTCGGAAATCGCCCCCTTCGCAGATCTCTCATGCCAGAGCGCTAAT 2045
DB 1355 TTGGCCAGATTTCCATGGCAACCGGTCTCCCTTAGCAGATCTGCACACTAAAGGCGATGGT 1414
QY 2046 TTGTGCCCTGGGTATGGAGCGAGATCTGGATTAATTTACTGCGCTTGTATATCGCTGGATT 2105
DB 1415 CACCGGATTTGAAACTGTCTCAGGACCTTGATGATCTTGCCATTTCTACTGTGCCACAGT 1474
QY 2106 ATGCGGAATTTGTTGTTGCTGGCCAAATTTCTGACGCTCAACACAGCGCAGGAGTAGT 2165
DB 1475 TCAAGCCATTGCTTTGGGACTCGCTTCATATAGAGGCCATGGAGGCGAGCGGCACCTC 1534
QY 2166 GAGTAAAAATATCGTTATTAGCGCGGTGCCGGCAGCATCCACTGTTACGACAAAATCT 2225
DB 1535 ATCAGTACTCTTTTCTATGTGGAGCCCTCAGCAAGATCCCTTTTGTGCAATGCA 1594
QY 2226 GGCAGATACCTCGGTTATTTCCGGTCAATACCACGCAATGCTGCGAACCCTTTTATTAGG 2285
DB 1595 TCGGACATTTACTGGCATGCTGTGCTGTGTCGCAAGAGTGGAGTCCGTTCTTGTGGG 1654
QY 2286 CTCGGCCATTTGTTGCTGTGCGCGGAATATTGACACCTTCTGTTGGCGAAGCGATGCA 2345
DB 1655 TGTCTGTTCTGGGTCGCTGTGCTGCTGAGGGATTTGCTGTTGACAGGAACAAATGGC 1714
QY 2346 ACAATTCACCCATGTGGAATAATATTATTCGCAAGAACGCTATCAATCTCTTCATCA 2405
DB 1715 AAAATGACAAAGTTGGGAAAGTTGTTGCCGAGACTACAGGATAAAAAATACTATGA 1774
QY 2406 TCGTCGATGAGGCCCTATAAGCAGTTG 2433
DB 1775 TAAGAAATACCAAGTATTCCTGAAGCTG 1802

RESULT 6
ABL26431
ID ABL26431 standard; DNA; 2602 BP.
XX
AC ABL26431;
XX
DT 26-MAR-2002 (first entry)

XX	Drosophila melanogaster genomic polynucleotide SEQ ID NO 30766.
DE	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical; gene; ds.
KW	Drosophila melanogaster.
OS	
XX	WO200171042-A2.
PN	27-SEP-2001.
PX	
PD	
XX	23-MAR-2001; 2001WO-US09231.
PF	
XX	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	(PEKE) PE CORP NY.
PA	
XX	Venter JC, Adams M, Li PWD, Myers EW;
PI	WPI; 2001-656860/75.
DR	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	Claim 1; SEQ ID NO 30766; 2lpp + Sequence Listing; English.
PS	The invention relates to an isolated nucleic acid detection reagent
XX	capable of detecting 1000 or more genes from drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 2602 BP; 705 A; 583 C; 669 G; 645 T; 0 other;
Query Match 5.3%; Score 211.6; DB 23; Length 2602;	
Best Local Similarity 48.8%; Pred. No. 5.3e-56;	
Matches 713; Conservative 0; Mismatches 714; Indels 33; Gaps 4;	
QY	878 TTATTGGTGTGATGGGATCAGGCAGTGTCGCCGGCGGGATTTTTCATCTCAACGGAT 937
DB	
DB	882 TTGTGGCGGTGGAGCTGGGCAGCGGCAGTGCAGGGCGGCCCTGGTCGCCCTGCCATGGGC 941
QY	938 CTCGTCTATCCATGCGCACAGAAAAATCACGACTACGCGCGCAGCGGAAGCCGCGTG 997
DB	
DB	942 GTGTCCTGGAGCAGCGGTGCAGACAGATCCAACAGTGGAACTCTGAGCCGGGCTACTACA 1001
QY	998 AACAGTCCAGCCAGGAGATCTGGCAGCGGCTGTCTTCATGTATTCGTAAATGCGCTCACTC 1057
DB	
DB	1002 ACCAGTCATCCGATAATATATGGCAGTCCATCTGCCAGGTGGTGAAGAAG-----TCA 1055
QY	1058 TGGCAGACGTTTGTGCACAAAGTGTGGAGCATCGGTTTTGTGATGCCACCTGTCTCTGG 1117
DB	
DB	1056 TTGCGCGCGTGCATTAAGTCAAAGTCAAGGCGATTGGCTTCGATGCCACCTGCTCTTTGG 1115
QY	1118 TGSTACTGGATAAAAACGGTGCATTCATTGGCTGTACGCCGGAAGGAGATCAAGACAA 1177
DB	
DB	1116 TGGTTCTGGCCCCCAGGATCGCCGCTGACGGGTGAGCAAGTCCGGTGAGCGGAGCAGA 1175
QY	1178 ATATCATTTGTGGATGGATCACCGCGCCACCGAACAGCAGAGCGAATCAATGCCACTC 1237
DB	
DB	1176 ACATAATCCTCTGATGGACCATCGTCCGAGAGAGGAACCCAGAGATCAATGCGCTTCA 1235
QY	1238 ACCATCGGGTCTGAACTAGCTCGGTGGTAAATTTTCGCTTGAATGAAACACCGGAAA 1297
DB	

RESULT 7

AAH15580

ID AAH15580 standard; cDNA; 1947 BP.

XX

AC AAH15580;

XX

XX 26-JUN-2001 (first entry)

XX

XX Human cDNA sequence SEQ ID NO:13883.

DE

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

KW

XX Homo sapiens.

OS

XX EP1074617-A2.

PN

XX 07-FEB-2001.

PD

XX 28-JUL-2000; 2000EP-0116126.

XX

XX 29-JUL-1999; 99JP-0248036.

PR

XX 27-AUG-1999; 99JP-0300253.

PR

XX 11-JAN-2000; 2000JP-0118776.

PR

XX 02-MAY-2000; 2000JP-0183767.

PR

XX 09-JUN-2000; 2000JP-0241899.

XX

XX (HELI-) HELIX RES INST.

PA

XX

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

PI

XX WPI; 2001-318749/34.

DR

XX

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT

XX full-length cDNAs defined in the specification, and for the detection

PT

XX and/or diagnosis of the abnormality of the proteins encoded by the

PT

XX full-length cDNAs -

PT

XX

XX Claim 8; SEQ ID 13883; 2537pp + CD ROM; English.

PS

XX

XX The present invention describes primer sets for synthesising 5602

CC

XX full-length cDNAs defined in the specification. Where a primer set

CC

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC

XX to the complementary strand of a polynucleotide which comprises one of

CC

XX the 5602 nucleotide sequences defined in the specification, where the

CC

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC

XX of an oligonucleotide comprising a sequence complementary to the

CC

XX complementary strand of a polynucleotide which comprises a 5'-end

CC

XX sequence and an oligonucleotide comprising a sequence complementary to a

CC

XX polynucleotide which comprises a 3'-end sequence, where the

CC

XX oligonucleotide comprises at least 15 nucleotides and the combination of

CC

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC

CC the specification. The primers can be used in antisense therapy and

CC

CC particularly full-length cDNAs. The primers are also useful for the

CC

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC

CC represent oligonucleotides, all of which are used in the exemplification

CC

CC of the present invention.

XX

XX Sequence 1947 BP; 500 A; 449 C; 515 G; 483 T; 0 other;

SQ

Query Match 4.6%; Score 183; DB 22; Length 1947;

Best Local Similarity 56.3%; Pred. No. 6.2e-47;

Matches 426; Conservative 0; Mismatches 295; Indels 36; Gaps 3;

QY 1160 AAGGAGATCGAAGCAAAATATCATGTGTGGATGGATCACCGCCGCCAACAGCAG 1219

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

||||| ||| | | | ||| ||||| ||||| ||| | | | |||||

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95993 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SQ Sequence 1044 BP; 262 A; 226 C; 274 G; 282 T; 0 other;

Query Match	3.1%	Score 124	DB 22	Length 1044
Best Local Similarity	47.8%	Pred. No. 3.1e-28		
Matches 392; Conservative	0	Mismatches 425	Indels 3	Gaps 1

Oy	1614	TGAGGGTTAATTGATTCCTACCGCTGTGGCATCGGTACGGTTGGCCGTATGAAAGGTGGAC	1673
Db	37	TGCGAGGAGTGAATTGGGGCAGATGTGAGAGGGCCGGCCCTCATCTGTGAGGGGGCCGACGT	96
Oy	1674	GCTAACAACATCTCGCGATATGTTTTCGGACATTTCTATCTACCAATGAGCATATTCACATTC	1733
Db	97	GAGCTCAACGGCTGGCTCTCATCTGTGGMAAGCTCTTCTTGTCAATGTGGGATCAGCAAGA	156
Oy	1734	TCCTCGTTGTACCGGGTGTCTGGGGGCCGTATTAACATGTGGATGGTTCAGGTCTGG	1793
Db	157	CCCCATTTTGTACACAGGCGCTGTGGGGGCGTTATTTCTCAGGCATAGTACCTGGGTTCTGG	216
Oy	1794	GTTAGTTGAAGGGGGCAAGGTCTCGGGAGACGCTATTGGACACAGTACTGATTTC	1853
Db	217	GCTAATATGAAGTGTGTCAAGCGCTTACTGGAAATTTGATAGACACATGTGTCAAGGCCA	276
Oy	1854	TCCGCGCTGTGAAGAACCTGCGCAAAATGGCACACGTGTAATCAGCCCGCTCCCGCTCG	1913
Db	277	TGCTGCTTTTCCAGAACCTACACAGTAAAGGGCACAGCAGATGCAGAGATATATGAT	335
Oy	1914	GCTTGCATGTGAATCTCTCCAAAAAGCGCGCACCATAGATGCTGTGCGCCCTGGCGAA	1973
Db	336	- -ATTGTAACACAGTACCCTGGATCTGATTAAGAAAGGCTCAGCCGTGGGTTCTCTTACGT	393
Oy	1974	AGGGCTACACGTTGGTCCGGGAATTTTCGGAATTCGCGCCCGCTTCCGAGATCCTCATGC	2033
Db	394	TGATTTACATGTTTGGCCAAATTTCCATGTGCMAACCGGTTCTCCCTTAGCAGATCTACACT	453
Oy	2034	CAGAGCGGTAAATTTGTGGCTGGGTATGAGACGAGATCTGATTAATTTACTGCGCTGTA	2093
Db	454	AAAGGGCATGGTCAACCGGATTTGAACGTGTCCAGGACCTTGATGATCTTGGCCATTCTCTA	513
Oy	2094	TATGCTGTGATTAATGCGGAATTTGGTATGTGTGCGCCAAATTTTCGACGCTCAAAAGC	2153
Db	514	CCTGGCCAGCTTCAAGCCATTGCTTTTGGGGAGCTCGCTTCATTTATGAGAGCCATGGAAGC	573
Oy	2154	GCAGGAGTATGTGATAAAAAATATCGTTAATTACGGCGGTGGCGGGCAGACATCACTGT	2213
Db	574	AGCAGGCGACTCATCTAGTACTCTTTTCCATGTGTGAGGCGCTCAGCAAGAAATCCCTTTT	633
Oy	2214	ACGACAAATTTCTGGCAGATACCTGCGGATATCCGGTGATTTACCAGCAGCAATGCTGGAAAC	2273
Db	634	TGTGCAAAATGATTCGGGACATTTACTGGCAATCGCTGTGGTCTGTGTCCAGAGGTGAGACTC	693
Oy	2274	GATTATTAAGGCTCGGCGCAATCTGTGGTGTGTCGCGCGGGAATATTTGCAACCTTGTGGG	2333
Db	694	CGTCTTGTGTGGTCTCTCTGTTTGGGTGCTGTGCTTACGGGATTTGCGTCTCTGTGACA	753
Oy	2334	CGAAGCATGCACAAATTTCAACCCATGTGATTAATTTATTCGCAAGAACGCTATCA	2393
Db	754	GGAAGCAATGGCAAAATTTGAGCAAAAGTCGGGAAGTTGTGTTCCGAGACTCAGAGATAA	813
Oy	2394	ATCTCTCATCATCTGTGATATAGGCTCTTAAAGCAGTTG	2433
Db	814	AAATACTATGATAGAAATACCAACTATTTCTCGAAGCTG	853

```

RESULT 10
AAI61235
ID AAI61235 standard; cDNA; 571 BP.

```

[illegible]

Query Match	2.88;	Score 112.2;	DB 22;	Length 571;
Best Local Similarity	57.68;	Pred. No. 1.2e-24;		
Matches 223;	Conservative 0;	Mismatches 158;	Indels 6;	Gaps 1;

[illegible]

Db 249 CCTGTGGCTTTTGAGACAGCAATTAAGATTGGAGCCCGCATTTCAACCAACCATGTA 308
 QY 999 ACAGTCCAGCAGAGATCTGGACAGCGGTCTGTTCATGTATTCGTATGGCTCAGCT 1058
 Db 309 GCAGTCCCTCCAGACATCTGGCGTGGCTGTGTTCACAAAGAAAGT-----TGT 362
 QY 1059 GGCAGACGTTTGTGCACAAAGTGTGGCAGCATCGCTTTTGTATGCGACCTGTTCTGT 1118
 Db 363 ACAAGGAGTGTATTAAACCAAAATCGAGAGACTTGGGTTTGTATGCGAGCTGTCTGT 422
 QY 1119 GGTACTGATTAACAGGATGATTCCTCTGTACGCCCGAGAGAGATGCAAGCAAAA 1178
 Db 423 TGTTTTGTATAGCAGTTTCAACCCCATTCACCAACGAGAGGGGATTCCTCATCGAAA 482
 QY 1179 TATCATTTGTGATGATGATCAACCGCGCACGACGAGACAGAGATCAATGCGCATCA 1238
 Db 483 CGTCACTCTGTGTGCTGGACCATTCGACAGTCAATCAATTAACAGATCAATGAGACCA 542
 QY 1239 CCATCCGCTGCTGAACACTACGTCGGTGG 1265
 Db 543 GCACAGTGTCTCCAGTACGTCGGGGG 569
 RESULT 11
 AAH04701
 ID AAH04701 standard; cDNA; 795 BP.
 AC AAH04701;
 DT 26-JUN-2001 (first entry)
 XX Human cDNA clone (5'-primer) SEQ ID NO:1536.
 DE Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 KW Homo sapiens.
 OS Homo sapiens.
 PN EPI074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 1; SEQ ID 1536; 2537pp + CD ROW; English.
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 SQ Sequence 795 BP; 174 A; 206 C; 224 G; 188 T; 3 other;
 Query Match 2.6%; Score 104.2; DB 22; Length 795;
 Best Local Similarity 57.7%; P-adj. No. 5.6e-22;
 Matches 241; Conservative 0; Mismatches 171; Indels 6; Gaps 3;
 QY 1160 AAGGAGATGCAAGCAAAATATCATTTGTGTGATGATCACCGCCACCAAGCAG 1219
 Db 380 AAGGGATTTCCATGGAACGATCATCATGTGCTGGACCATGACATCACTCAAGTTA 439
 QY 1220 AGCGAATCAATGCCACATCCGATCGGTCTGAACCTACGTCGGTGAATTTGCGCTG 1279
 Db 440 ACAGGATCAATGAGACCAAGCAGATGCTCCAGTACGTCGGGGGGTGTGTGTG 499
 QY 1280 AATGGAACACCGAAATTTCTGTGCTGGAAGAAATATGCAAGAT---CTACGAAAC 1336
 Db 500 AATGCAAGCCCGGAAACCTTGTGTGCTGGAAGAAACCTTGTGGAAGATTTGCTGGATA 559
 QY 1337 GTGCCGACAAATTTTGCATGCGCCGATTTTCTGACCTGGGCGGCTACCGGTGATTAG 1396
 Db 560 AGCGCGGACATTTCTTTGATCTCCCGGACTTCTTTCGTGGAAGCAACAGGTGTCAAG 619
 QY 1397 CGCGTTCAATGATGACATGTTTACCTGTAATGAGACGTGCGCATGAAATTCGCTGG 1456
 Db 620 CACGGGCTCTGCTGCTCCCTGGGTGTAA--GTGACATATTCACACAGAAAGCTGG 677
 QY 1457 ATCCAGATTATTTCCGACACATGCGCTTGCAGAGTTAGCGGATTAAGATTTTATCCGA 1516
 Db 678 ACGACAGTTTCTGGAATAATGATGNTTGGAAAGCTTTGTGTCAAG-ATAATTCACACAAA 736
 QY 1517 TTGTCATCATATTTGTTTCTCCGACACACCTTGTGAAATGTTTACAGCAGCAAGC 1574
 Db 737 TAGGAACCAAGTCTTACTTGTGAGACTTCTCTTGGAAATGGCTTACACCAAGANCC 794
 RESULT 12
 ID ABN60111 standard; cDNA; 677 BP.
 AC ABN60111;
 DT 28-JUN-2002 (first entry)
 XX
 DE Human cancer related polynucleotide SEQ ID NO 78.
 KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
 KW gene therapy; cancer; tumour; gene; ss.
 OS Homo sapiens.
 PN WO200214500-A2;
 PD 21-FEB-2002.
 PF 16-AUG-2001; 2001WO-US25840.
 PR 16-AUG-2000; 2000US-226326P.
 PA (CHIR) CHIRON CORP.
 PA (HXSE-) HXSEQ INC.

XX Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
 PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
 XX WPI: 2002-241905/29.
 XX
 PT New nucleic acid for producing a polypeptide, detecting differentially
 PT expressed genes correlated with a cancerous state of a mammalian cell,
 PT and inhibiting tumor growth -
 XX
 PS Claim 1: SEQ ID NO 78; 883bp + Sequence Listing: English.
 CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
 CC with cytosolic activity. The polynucleotide is used to produce a
 CC polypeptide, to detect differentially expressed genes correlated with a
 CC cancerous state of a mammalian cell and to inhibit tumor growth. The
 CC polynucleotide is used as a probe in mapping and tissue profiling. The
 CC encoded polypeptide and antibodies to the polypeptide can also be used
 CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
 CC gene therapy.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 677 BP; 157 A; 164 C; 184 G; 172 T; 0 other;

Query Match 2.3%; Score 90.4; DB 24; Length 677;

Best Local Similarity 49.4%; Pred. No. 1.3e-17;
 Matches 321; Conservative 0; Mismatches 296; Indels 33; Gaps 2;

QY 1525 CATATGTTTCCCGGACACCTGTGGAATGTTTAAACACACAGCCGGCAGAG 1584
 DB 1 CAAGTCTACCTCTGAGAGCTTCTTGGAAATGGCTCACACAGAGCAGCAAGAGAC 60
 QY 1585 ATGGAGTATCTCCCGGACACCTGTGCTGTAGGGTTATGATGATCTCAGCTGTGGC 1644
 DB 61 CTGGGCTTCTCCCTGGGATGGCGTGGCAGCTTCACTATGATGCCATGACGAGGA 120
 QY 1645 ATCGGTACGCTTGGGCTAGAGGTGAGCGCTAACACATCTCGCGTATG----- 1693
 DB 121 CTAGAGTATGGGGGAGGTGAGAGGGGACAGGCGCTCTGTAGGGGACAGCAGTG 180
 QY 1694 -----TTTCCGACCTTTCATGACCATGAGCATCTACCACTTC 1734
 DB 181 ACGTACGCGCTGCTCATCTGTGAGACGCTTCTTGTGACATGGGATCACAACAGAC 240
 QY 1735 CCCTGCTTTTACCGGGGTCTGGGGCCGTATTACAGTCCAGTGTCCAGGTCTGTGG 1794
 DB 241 CCGATTTTGTACAGGCGCTGGGGCCCTTATTCTCAGCCATGGTACTGGGTTCTGG 300
 QY 1795 TTAGTTGAAGGGGCGCAAGGTCTGGGAGACAGCTATTGACCACTTGTGATTTCCAT 1854
 DB 301 CTGAATGAAGGTGTCAGAGCGTTACTGGAATAATGATAGCACATGTGACAAGGCAT 360
 QY 1855 CCGGCTGTGAAGAGCTCGCGAATGCGACACAGTGTGATGATGAGCCCTCCCGCTGG 1914
 DB 361 GCTGCTTTTCCAGAACTTACAGTAAGGACACAGCCAGATGCCAGATATATATGAT-- 418
 QY 1915 CTTCGTGATGATCTTCGCAAAAACGCGCAACATCATGATCTCTCCCGCGGAGAA 1974
 DB 419 -ATTGACAGTACCTGATCTGATTAAGAAAGCTCAGCCTGTGGTTTCTTACTGTT 477
 QY 1975 GGGCTACAGCTGGTGGCGAATTTCTCGAATATGCGCCCTTCGCAATCTCATGCC 2034
 DB 478 GATTTACATGTTTGGCAGATTTCCATGCAACCGGCTCCCTTACAGATCTGACACTA 537
 QY 2035 AAGAGCGTATTTTGGCTGGGTATGAGACGAGATCTGATATTTACTCGCTTGTAT 2094
 DB 538 AAGGCGATGATACCGGATTTGAACGTCTCAGGACCTTGATATCTTGCATTTCTCTAC 597
 QY 2095 ATCGGTGATTTATGCGGAATTTGTTATGTCGCGCAATTCGACGG 2144
 DB 598 CTGGCCACAGTTCAAGCCATGCTTTGGGGACTCGCTTCAATATATAGAGC 647

RESULT 13
 AAC02752
 ID AAC02752 standard; cDNA; 409 BP.

XX AAC02752;
 AC 06-OCT-2000 (first entry)
 DT
 XX

DE Human secreted protein 5' EST, SEQ ID NO: 2750.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PA Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.
 DR P-PSDB: AAC02746.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 2750; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 CC

SO Sequence 409 BP; 103 A; 89 C; 115 G; 102 T; 0 other;

Query Match 2.2%; Score 86.2; DB 21; Length 409;

Best Local Similarity 56.8%; Pred. No. 2e-16;
 Matches 179; Conservative 0; Mismatches 133; Indels 3; Gaps 1;

QY 1329 CTACGACGTCGGGCAATTTTGTGATCGGCGGATTTTCTGACCTGGGCGCTACCGG 1388
 DB 93 CTGGGATTAAGGCGGACATTTCTTGTATCTCCGGAATCTTATCTGGAAGCAACAG 152
 QY 1389 TGATTTAGCGCGTTCAGTATGACTGTACCTTAATGAGCGTGGCTGACATGAAA 1448
 DB 153 TGTACAGCAGCGTCTCTGCTCCCTGGTGTAAAGTGACATATTCAGCA---GAGA 209
 QY 1449 TCCTGGGATCCAGATTAATTTCCGACATTTGCTTGCAGACTTGAAGGATGAAGATT 1508
 DB 210 AGGCTGAGACAGATTTCTGCAAAATGATTTGGTTGGAAGACTTGTGACAGATAATTA 269
 QY 1509 TATTCGATTTGATCATATTTCTTCTCCGGAACACTGTGGAATATGTTAACACG 1568

Db 270 CAGCAAAATAGCAACCAAGTGTACTCTCTGGAGCTTCTGTGAAATAGGCTCAGACC 329
 OY 1569 ACAACCGCGGAGATGATGATATCTCCCGGACACACTGTCGTAGGTTAATGCA 1628
 Db 330 AGAGGACAGCAAGACCTTGCTCTCCCTGGGATGTCGGGAGGAGCTTCACTCATTTGA 389
 OY 1629 TGCTCAGCGCTGTGG 1643
 Db 390 TGCCCATGCGAGGAGG 404

 RESULT 14
 AAH07400
 ID AAH07400 standard; cDNA; 613 BP.
 AC AAH07400;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:4235.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Mshikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 1; SEQ ID 4235; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX SQ Sequence 613 BP; 146 A; 140 C; 164 G; 158 T; 5 other:
 Query Match 2.1%; Score 82; DB 22; Length 613;
 Best Local Similarity, 48.4%; Pred. No. 5,9e-15;
 Matches 257; Conservative 0; Mismatches 271; Indels 3; Gaps 1;

 OY 1614 TGTAGGTTAATTGATGCTCAGCGTGTGTCAGTTCAGTGTGGGCTGAAGGTGAGC 1673
 Db 37 TGCAGAGTGATTTGGCGGAGATGTGAGAGGCGACGCCCTCATCTGTGAGGGCGGCCACT 96
 OY 1674 GCTGAACAATCTCGGGTATGTTTGGCACTTCTTCATGCACACATGACATTCACACTTC 1733
 Db 97 GACGTACAGCGCTGGCTGTGATCTGTGGAACGCTTCTTGTCAATGAGGGATCAGCAAGA 156
 OY 1734 TCCCTCGTTTGTACCGGGGTGTGCGGGCCGTATTACACTGCGATGTTCACAGTCTGTG 1793
 Db 157 CCCGATTTTGTACCAAGGCGCTGGGGGCTTATTTCTCAGCCATGTGCTGCGTTCG 216
 OY 1794 GTTAGTGAAGGCGGCAAGTGTGCGGAGCAGCTATTTGACGACGCTTGTGATTTCGA 1853
 Db 217 GCTGATGAGAGTGTGCAAGCGTTACTGGAATTTGATGACACATGCTACAGGCCA 276
 OY 1854 TCCGGCTGTTGAAGAAGCTCGCGAAATGCGCACAGCTGTGAATCAGCCCTCCGCTGTG 1913
 Db 277 TGCTGCTTTCCAGAACTCAAGTAAAGGCCACAGCCACATGCGAGATATATATGACATA 336
 OY 1914 GCTTGCTATGCAATCTTCGAAAAAGCGGCACACCATCAGATGCTGTGCGCTGGCGAA 1973
 Db 337 TTTGAACAGTC--ACCTGATCTGATTAAGAGGCTCAGCCGTGGGCTTCTTACTGT 393
 OY 1974 AGGCGTACAGCGTGGCGGAATTTCTGGAATATGCGGCGCCCTGCGAGATCTCATGTC 2033
 Db 394 TGAATTACATGTTTGGCCAGATTTCCATGCAACCGGTCTCCCTTAGCAGATCTGACACT 453
 OY 2034 CAGAGCGTAATTTTGGCTGGGTATGAGGCGAGATCTGATTAATTTACTGCGCTTGTGA 2093
 Db 454 AAGGCGATCGTCCACCGGATTAAGACTGTCTCAGAGCCTGTGATGATCTTGCCATCTCTGA 513
 OY 2094 TATCGCTGATTAATGCGGATTTGTTATGCTGTGCGCCCAATTTCTGACGC 2144
 Db 514 CCTGGCCACAGTTCAAGCCATTGCTTGGGGAGCTGCTTCANTATATAGAAGC 564

 RESULT 15
 AAF11917
 ID AAF11917 standard; cDNA; 640 BP.
 AC AAF11917;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:4440.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Aspergillus oryzae;
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PR 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX WPI; 2000-594572/56.
DR
XX
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX
XX Claim 88; Page 1897; 3161pp; English.
PS
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from
CC *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.
XX
XX
SQ Sequence 640 bp; 166 A; 125 C; 183 G; 166 T; 0 other;

Query Match 2.0%; Score 78.4; DB 21; Length 640;
Best Local Similarity 53.4%; Pred. No. 8.6e-14;
Matches 191; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

QY 1073 CACAAAGTGGCGAGCATGGTTTGGATGCGCACTGTCCTGCGGTACT-----GG 1126
DB 15 CAGAAAACCGTACTGGGATTTGGCTTGGATGCAACATGTTCTGCTTCCACAG 74

QY 1127 ATAAAAACGGTGATCCATTGCTGTCTAGCCCGGAGGAGATGCCAAATATCATTTG 1186
DB 75 TTACCGCAGAACCTATATCTGTACGGGACCAATTTCGATTCGATCGCATGTCTATCC 134

QY 1187 TGTGATGATCTACCGCGCCACCGAACAGCAGAGCAATCAATGCCACTCACCACCGG 1246
DB 135 TTTGGGTAGATCATCGCCCGTGGAGAGACGGAAGTTAATCTACCAATCATTAAT 194

QY 1247 TGCTGAACCTAGCTGGGTGTAATTTGCGCTGAATGGAACACCGAAATTTCTGCGC 1306
DB 195 TGCTTGTTAGCTGGTGGAGAAATGTCATTGAATGAGATTCCCAAGGTCCTGTGGC 254

QY 1307 TGAAGAAATATATCCAGAGATCTACGAACGTCCGACAAATTTTTCGATCTGGCGATT 1366
DB 255 TGAATAATCATATGCCCAAGATCTGTTCAGCAGTCGCAAGTCTATGATCTGGCGATG 314

QY 1367 TTCTGACCTGGCGGCTACCGGTGATTTAGCGGTTCTAGTATGCACCTGTTACCTGTAA 1424
DB 315 CACTGACCATATTCGCAACTGCAATGAGAAAGAAATTTCTAGCGTTGTTGTAA 372

Search completed: March 21, 2003, 20:48:59
Job time : 910.004 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 09:44:37 ; Search time 143.991 Seconds
(without alignments)
8457.593 Million cell updates/sec

Title: US-09-802-208B-2

Sequence: 1 atcgattgagcagtttgct.....tcacgcgctcagccatcgat 3971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```
IssueParents_NA: *
1: /cgn2_6/prodata/1/lna/5A_COMB.seq.*
2: /cgn2_6/prodata/1/lna/5B_COMB.seq.*
3: /cgn2_6/prodata/1/lna/6A_COMB.seq.*
4: /cgn2_6/prodata/1/lna/6B_COMB.seq.*
5: /cgn2_6/prodata/1/lna/POTUS_COMB.seq.*
6: /cgn2_6/prodata/1/lna/backfillseq.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	311	7.8	1401	4	US-09-134-001C-183	Sequence 183, App
2	69.6	1.8	696	4	US-09-134-001C-1560	Sequence 1560, App
3	68.8	1.7	744	4	US-09-385-028-22	Sequence 22, App
4	68.8	1.7	11604	4	US-09-385-028-13	Sequence 13, App
5	68.8	1.7	15079	4	US-09-385-028-1	Sequence 1, App
6	67	1.7	6133	4	US-09-453-702B-15	Sequence 15, App
7	57.2	1.4	5395	4	US-09-421-017B-383	Sequence 383, App
8	56	1.4	1155	3	US-08-793-035-2	Sequence 2, App
9	56	1.4	1185	3	US-08-793-035-1	Sequence 1, App
10	53.8	1.4	1512	1	US-08-594-808B-6	Sequence 6, App
11	52.2	1.3	1670	3	US-09-026-882B-1	Sequence 1, App
12	51.8	1.3	634	4	US-09-296-284-67	Sequence 67, App
13	51.8	1.3	684	4	US-09-296-284-66	Sequence 66, App
14	51.8	1.3	734	4	US-09-296-284-65	Sequence 65, App
15	51.8	1.3	784	4	US-09-296-284-63	Sequence 63, App
16	51.8	1.3	784	4	US-09-296-284-64	Sequence 64, App
17	51.8	1.3	834	4	US-09-296-284-62	Sequence 62, App
18	51.8	1.3	884	4	US-09-296-284-61	Sequence 61, App
19	51.8	1.3	934	4	US-09-296-284-60	Sequence 60, App
20	51.8	1.3	984	4	US-09-296-284-59	Sequence 59, App
21	51.8	1.3	1034	4	US-09-296-284-58	Sequence 58, App
22	51.8	1.3	1084	4	US-09-296-284-57	Sequence 57, App
23	51.8	1.3	1134	4	US-09-296-284-56	Sequence 56, App
24	51.8	1.3	1184	4	US-09-296-284-55	Sequence 55, App
25	51.8	1.3	1234	4	US-09-296-284-54	Sequence 54, App
26	51.8	1.3	1284	4	US-09-296-284-53	Sequence 53, App
27	51.8	1.3	1334	4	US-09-296-284-52	Sequence 52, App

28	51.8	1.3	1383	4	US-09-326-884-51	Sequence 51, Appl
29	51.8	1.3	2700	4	US-09-396-284-51	Sequence 51, Appl
30	50	1.3	1014	2	US-08-673-1904-13	Sequence 13, Appl
31	49.4	1.2	7850	4	US-09-134-001C-1388	Sequence 1388, Appl
32	45.6	1.2	763	4	US-09-221-017B-723	Sequence 723, Appl
33	43.2	1.1	5338	4	US-08-961-527-50	Sequence 50, Appl
34	42.4	1.1	31728	4	US-09-453-702B-64	Sequence 64, Appl
35	42.2	1.1	3058	4	US-09-221-017B-757	Sequence 757, Appl
36	41.8	1.1	741	3	US-09-338-481-1	Sequence 1, Appl
37	41.8	1.1	741	4	US-09-572-810A-1	Sequence 1, Appl
38	41.8	1.1	7218	1	US-08-232-463-14	Sequence 14, Appl
39	41	1.0	732	4	US-09-339-052-1	Sequence 1, Appl
40	40.4	1.0	835	4	US-09-134-001C-1594	Sequence 1594, Appl
41	39.4	1.0	19702	4	US-08-961-527-78	Sequence 78, Appl
42	39.9	1.0	1436	4	US-08-858-207A-78	Sequence 78, Appl
43	38.6	1.0	1086	1	US-08-664-400-1	Sequence 1, Appl
44	38.6	1.0	1086	1	US-08-674-273A-1	Sequence 1, Appl
45	38.6	1.0	1086	4	US-09-123-286-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-183

GENERAL INFORMATION:

: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

```

: NAME/KEY : unsure
: LOCATION : (157)
: OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-134 -001C-183

```

[illegible]


```

; STREET: The Jenifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELE: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-385-028-22

```

Query Match 1.7%; Score 68.8; DB 4; Length 744;

Best Local Similarity 47.4%; Pred. No. 1e-11; Matches 279; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

```

QY 120 ATGATATCTCCCTTAATGCAAGTTGCGAGTATCATGCGCTGCTGAGTATTGCG 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 744 ATGCCATCCGCACTCCAGGGAAAGTGGCGTCATCAGCGGCGGAGCTCGGCAATCGCG 685
QY 180 CTGCAATGTGCAAAAAGCTGCTCGATGCAAGCAAGTAAGTATTGATTGACCGGAA 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 684 GAGGCGACGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625
QY 240 GCGGCAAACTGCACAAGTTGTCTGAGT-----AGCGCAAAACGGTACGGG 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 624 GTGGAAGAACTGCGCGCTGCGGTGACGAGTGAACCGCGCGCGCGCGCGCGCGCGCG 565
QY 291 CTGCAACTGATCTCTTCAATATATGCAAGTGCATATGCTGCGGCAATTATCGAA 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 564 CTGCAACTGAGCTGCCCAACCGGCGAGGGGTGACGCGCGCGCTCCCTTCACCGTCGAG 505
QY 351 CTGGCGGGTGGGCTGATATTTTTCATGCCAATGCAAGCGCTTATATTTGGCGGCCAGTG 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 504 GGGCTGGGGGGGCTGACATCTCGTCAACAAGCGCGGATCATGTGCTGGCGCGCGG 445
QY 411 GCTGAAGTGATTCAGATGCTGCGGATGCTGTGTTAAATCTGAATATTAATCGCGGCTTT 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 GAGGAGCGCGACACCACTGACCGCGATGATGACCAACATCTCTGGGCGCTGATG 385
QY 471 GCGTGTGTCGAGTCTGCGCGCATATGATGCGCAAGTGGCGGCGCATATATTTT 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 TACATGACCCGGGCGCGCTTCCCATCTGCT---GCGCAGCAAGGGCACCAGGTCGAG 328
QY 531 ACCAGTTCCATCGCGGGCTGCTCGGTATCTGGGAACCGATACACCGCGTCAAAA 590
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 327 ATGTCTCTGATCGCGGGCGGGGTAACGTCGCAACGCGCGCGTCTACAGAGCGCAAGAG 268
QY 591 TTTCGGCTTCAGGATCTGTACACACTACCGCGCGCGAGTTTCTCAATATGCGCGT 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 TTCGGTGTGAACGCTTTCAGCAGACGCTGCGCCACAGAGGATCACCGAGCGCGGGGTGCGG 208

```

```

QY 651 GTGGGCGGTGCTGCGCAGACAGTACTGCGCTGTTGATGAC 698
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 207 GTCTGCTGATGAGCGCGGCGACACCGCAGCAGGAGCTCGCGGCCAC 160

```

RESULT 4

US-09-385-028-13/c
Sequence 13, Application US/09385028
Patent No. 6232106

GENERAL INFORMATION:
APPLICANT: Susan E. Jensen

APPLICANT: Kwame A. Aidoo

APPLICANT: Ashish S. Paradkar

TITLE OF INVENTION: DNA sequence Encoding Enzymes of Clavulanic
Patent No. 6232106

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSER: JACOBSON, PRICE, HOLMAN & STERN, PLLC

STREET: The Jenifer Building, 400 Seventh Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/385,028

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/790,462

FILING DATE: 29-JAN-1997

ATTORNEY/AGENT INFORMATION:

NAME: D. Douglas Price

REGISTRATION NUMBER: 24,514

REFERENCE/DOCKET NUMBER: 1418/P57452US2

TELEPHONE: (202) 638-6666

TELEFAX: (202) 39305350

IDEA UR

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 11604 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-385-028-13

Query Match 1.7%; Score 68.8; DB 4; Length 11604;

Best Local Similarity 47.4%; Pred. No. 6.9e-11; Matches 279; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

```

QY 120 ATGATATCTCCCTTAATGCAAGTTGCGAGTATCATGCGCTGCTGAGTATTGCG 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1133 ATGCCATCCGCACTCCAGGGAAAGTGGCGTCATCAGGCGCGGAGCTCGGCAATCGCG 11274
QY 180 CTGCAATGTGCAAAAAGCTGCTCGATGCAAGCAAGTAAGTATTGATTGACCGGAA 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11273 GAGGCGACGCGCGCGCTGCGGTGACGAGGGCGCGCGCTGCGCATTCGCGCGCGCG 11214
QY 240 GCGGCAAACTGCACAAGTTGTCTGAGT-----AGCGCAAAACGGTACGGG 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11213 GTGGAAGAACTGCGCGCTGCGGTGACGAGTGAACCGCGCGCGGCGGCAAGTCTGTC 11154
QY 291 CTGCAACTGATCTCTTCAATATATGCAAGTGCATATGCTGCGGCAATTATCGAA 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11153 CTGCAACTGAGCTGCCCAACCGGCGAGGGGTGACGCGCGCGCTGCTCCACCGTCGAG 11094

```

```

QY 351 CTGGCGGGTGGGCTGATATTTTTCATGCCAATGACAGCGCTTATATTTGGCCCAATG 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11093 GCGGTGGGGGGCGCTGACATCTCTGTCACACAGCGCGGATCATGCTGCGGCCGTTG 11034
QY 411 GCTGAAGTATCCAGATGCTGTGGATGCTGTGTTAAATCTGAATATTAATGCGGCTTT 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11033 GAGGACCCCAACACACAGCATGACCCCGATGATCAGACACATCTCTGCGGCTGATG 10974
QY 471 CGGTGTCGTGAGTCCGCGCATATGATGCGAGAGGTCGAGAGTGGCGGATATATTTT 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10973 TACATGACCCGCGGCGCTTCCCATCTGCT--GCGCAGCAAGGCGACCGTGTGAC 10917
QY 531 ACCAGTTCATCGCGGGCGTCTGTCGGTTATCTGGAGACCGATCTACACCGCTCCAA 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10916 ATGCTCTGATCGGGGCGGGTGACGTCGCGCAAGCGGCGCTCTACAGGCCACGAG 10857
QY 591 TTTGCCCTTACGATCTCTGATACACATACCGCGCCGCAAGTTTCTCAATATGCGTGGT 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10856 TTGCGGTGTGAACGGTTTCAGACGACGCTGCGCAGAGGTACACCGAGCGCGGGTGG 10797
QY 651 GTGGGTGCGGTGCTGCCAGACAGTAGTACGTCGCTGCTGTGATGAC 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10796 GTGCTGCTATCAGCGCGGCGACACGACGAGAGCTGCGGCGCCAC 10749

```

```

RESULT 5
US-09-385-028-1/c
; Sequence 1, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A Aidoo
; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, P/LC
; STREET: The Jennifer Building, 400 Seventh Street, N.W.
; City: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15079 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

```

```

; ORIGINAL SOURCE:
; ORGANISM: Streptomyces clavuligerus
; US-09-385-028-1
Query Match 1.7%; Score 68.8; DB 4; Length 15079;
Best Local Similarity 47.4%; Pred. No. 8.3e-11;
Matches 279; Conservative 0; Mismatches 297; Indels 12; Gaps 2;

```

```

QY 120 ATGAATACCTCCCTTATGCAAAAGTGTGACGATCTGCTGCGTCCGATGATGCG 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13365 ATGCAATCCGCACTCCAGGGGAAAGTCCGCTCTATCAGCGGCGGAGCTGGCATGCG 13306
QY 180 CTGCAATGTCCAAAGAGCTGCTGATGACAGAGCAAGATGATGATGATGACCGGAA 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13305 GAGGACCAAGCGCGCGCTGCGCGCGAGGCGCGCGCTGCGCATGCGCGCGCG 13246
QY 240 GCGGCAAACTGCAAGATTTGCTGAGT-----AGCGCAAAAGCGGTACGGC 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13245 GTGAGAAAGCTGCGCGCGCTGCGGAGACGACGACCGCGCGCGCGCGCGCGCG 13186
QY 291 CTGCAATCTGATCTCTTCAATTAATCAGCAAGTCAATGCTGCGGCAATTAATGAA 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13185 CTGCAATCTGATCTCTGCGGAGACCGCGGAGGAGGAGCGCGCGCGCGCGCGCG 13126
QY 351 CTGGCGGGTGGGCTGATATTTTTCATGCCAATGACAGCGCTTATATTTGGCGGCCAGTG 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13125 GCGCTGGCGCGCGCTGCGATCTCTGTCACACAGCGCGGATCATGCTGCTGCGCGCGTG 13066
QY 411 GCTGAAGTGTATCCAGATGCTGCGATGCTGTTAATCTGAATATTAATGCGCGCTTT 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13065 GAGGACCGCGGACACACCGCATGACCGCGGATGATGACACCAATCTCTGCGGCTATG 13006
QY 471 CGGTGTCGTGAGTCTGCTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 530
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13005 TACATGACCGCGGCGCGCTTCCCATCTGCT--GCGCAGCAAGGCGACCGTGTGAC 12949
QY 531 ACCAGTTCATCGCGGGCGTCTGCGGTTATCTGGAGACCGATCTACACCGCTCCAA 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12948 ATGCTCTGATCTGCGGGCGCGGATGACGTCGCGGCGCGCGCTCTACAGGCCACGAG 12889
QY 591 TTTGCCCTTACGATCTCTGATACACATACCGCGCGGATGATGATGATGATGATGATGAT 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12888 TTGCGGTGTGAACGGTTTCAGACGACGCTGCGCGAGAGGTACACGAGCGCGGGTGG 12829
QY 651 GTGGGTGCGGTGCTGCCAGACAGTAGTACGTCGCTGCTGTGATGAC 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12828 GTGCTGCTATCAGCGCGGCGACACGACGAGAGCTGCGGCGCCAC 12781

```

```

RESULT 6
US-09-453-702B-15
; Sequence 15, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatner, Frederick R.
; Buriand, Nicole T.
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plunkett Street
; City: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0

```

```

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/453,702B
  FILING DATE: 03-DEC-1999
  CLASSIFICATION: <UNKNOWN>
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 60/110,955
    FILING DATE: 04-DEC-1998
  ATTORNEY/AGENT INFORMATION:
    NAME: Seay, Nicholas J.
    REGISTRATION NUMBER: 27386
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (608) 251-5000
    TELEFAX: (608) 251-9166
  INFORMATION FOR SEQ ID NO: 15:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 6133
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-453-702B-15

```

```

Query Match      -1.7%; Score 67; DB 4; Length 6133;
Best Local Similarity 47.7%; Pred. No. 1.7e-10;
Matches 236; Conservative 0; Mismatches 250; Indels 9; Gaps 1;

```

```

QY 119 TATGAACTACCTCCCTTAATGCAAGTTCGACCTATCATCTGGCGCTGCGTCAAGTATTGG 178
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 644 TTTGAGGAGAGCAATGATGATAGCTATATTAATTCAGCGGCGATCAAGTGATATGG 703
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 CCTGAATGTGCAAAACGCTGCTGCATGCAGAGCAAGAGTATGATGATGACCGGGA 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 704 GGAAGTATTTGCCAGAGAGCTTGGAATGACAGGTGCAAAAGTTTACTGGAGCACGCG 763
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 239 AGGCACAACACTGCACAAAGATTGTGCTGAGTT-----AGCGGAAAACGCGTACGC 289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 764 AGTTAGCGTATCGAAGCATTCGAACGGAATCTGCCGCGAGGAGAAATTGCTAAAGC 823
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 GGTGAACCTCGATCTCTTCAATATCAGCAAGTGCATACATGCTGGCGGACATTTATCA 349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 824 GCGGAGATTGATGTCACAGACGACAGTCATGCGCGGATTCGTCGCAAGCAGCGCTGA 883
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 350 ACTGGCGGTGGGCTGATATTTTCATGCGCAATGCGAGGCGCTTATATTGGGCGCCACT 409
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 884 TAGCTGGGGCGAGTTGATGTTCTTATCATATATGCGGCGTTATGCCCTTTCACCGCT 943
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410 GCGTGAAGTGAATCCAGATGTCGTGGATCGTGTGTTAAATCTGAATATTAATGCGCGCT 469
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 944 TGCACAGGAAACAAAGATGATGGGCGCTCAGTATGACGTGAATATCAAAAGGTGTACT 1003
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 470 TCGCTGTGCTCGTCAAGTCTCCGCAATATGATTGGCGCAGAGGTGGGCGGATATAATTTT 529
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1004 GTGGGGAATTTGGCGCTGACTTCCGCTGATGGAACACACAGGCTTCGCGTCAATATCAA 1063
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 TACCAATTCACGCGGGGCGGTCCGCTTATCTGGGAACGAGATTAACGCGCGCA 589
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1064 TCTTGTTCTATTGGTGGCCCTTCTGTTGTGCCACAGCGCAGTCTATTGTGCATCAAA 1123
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 590 ATTTGCCCTTACGC 604
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1124 ATTTGACATGAGGCG 1138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 7
US-09-221-017B-383/C
; Sequence 383, Application us/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120

```

```

CORRESPONDENCE ADDRESS:
  ADDRESSEE: MORRISON & FOERSTER
  STREET: 755 PAGE MILL ROAD
  CITY: Palo Alto
  STATE: CA
  COUNTRY: USA
  ZIP: 94304-1018
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: Windows
  SOFTWARE: FASTSEQ for Windows Version 2.0b
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/221,017B
    FILING DATE: 23-DEC-1998
  CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PP1182
      FILING DATE: 31-DEC-1997
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PP1546
      FILING DATE: 30-JAN-1998
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PP2911
      FILING DATE: 09-APR-1998
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: PCT/AU98/01023
      FILING DATE: 10-DEC-1998
  ATTORNEY/AGENT INFORMATION:
    NAME: Montoy, Gladys H
    REGISTRATION NUMBER: 32,430
    REFERENCE/DOCKET NUMBER: 27340-20021.00
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 650-813-5600
      TELEFAX: 650-494-0792
    TELEX: 706141
  INFORMATION FOR SEQ ID NO: 383:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 5395 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: circular
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ANTI-SENSE: UNKNOWN
    ORIGINAL SOURCE:
      ORGANISM: PORPHYROMONAS GINGIVALIS
      FEATURE:
        NAME/KEY: misc.feature
        LOCATION: 1...5395
US-09-221-017B-383

```

```

Query Match      1.4%; Score 57.2; DB 4; Length 5395;
Best Local Similarity 47.1%; Pred. No. 2.3e-07;
Matches 176; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

QY 320 ACTCATATACATCGTGGCGGCGCATATATGAACTGGGCGGTGGATATTTTTCATGC 379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1370 ACCACATACAGTGTGGAGACAGATCAAGCTCAGTTCGCTGAGATGATATTTTGGTAA 1311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 380 CAATGAGGCGCTTATATTTGCGGCGCCAGTGCCTGAAGGTGATCCAGATGCTGGAGTCG 439
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1310 CAATGCGGTATTTACTCGGAGATGGCGCTTATGATGGTATGACCGAGCAACATGGAGTGC 1251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 440 TGTGTTAATCTGATATTAATATGCGCGTTTGGCTGTGCTGCGTCAAGTCTCGCCGATAT 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1250 CGTGATCAACGTTAATCTGAAGTCAAGCTTCAACATGATCCATGCCGTACTCCATCAT 1191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 500 GATTGGCAGAGTGGGCGGATATATTTTTCAGATTCCATGCGGCGGCGTTCGCGGT 559
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1190 GATGGACAGCGCCACCGCGAGTATCATATATGCTTGTGTGTGTGTGTGTGTGTGTGTGT 1131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 560 TATCTGGGAACGATCTAACCCGCGTCCAAATTTGGCGTTGAGGCAATTCGTACACACTAC 619
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 1130 TCGAGGCAATCCAACTACACGCTTCTTAAGCCGGATATGCGTTGGCCAAAGACTGT 1071
QY 620 CCGCCCGGAGTTTCTCAATATGCGGTGCGTGGCGGTGCGGACAGACTGT 679
Db 1070 CGCCAAAGAGCTGGCGCTGACGTGGCGTGGCGTGGCGCAATGCAATGCAACCGGTTTATCAT 1011
QY 680 CACTGCCCTGCTTG 693
Db 1010 CACCGATATGACTG 997

RESULT 8
US-08-793-035-2
; Sequence 2, Application US/08793035
; Patent No. 6011201
; GENERAL INFORMATION:
; APPLICANT: Slabas, Antoni R.
; APPLICANT: White, Andrew
; APPLICANT: Chase, Dianne
; APPLICANT: Elborough, Keiran
; APPLICANT: Fentem, Phillip A.
; TITLE OF INVENTION: B-ketocyl ACP Reductase Genes from
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,035
; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9414622.2
; FILING DATE: 20-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB95/01678
; FILING DATE: 17-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:132
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713.787.1400
; TELEFAX: 713.787.1440
; INFORMATION FOR SEQ. ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-793-035-2

Query Match 1.4%; Score 56; DB 3; Length 1155;
Best Local Similarity 43.1%; Pred. No. 1.9e-07;
Matches 269; Conservative 0; Mismatches 355; Indels 0; Gaps 0;

QY 109 CTGTGCCCTCTATGAATCTCCCTTAAGCAAAAGTTCAGCTATCATCGCGCTGCGT 168
Db 209 CTGTTCGAAAGTGAAGTCTCCGCTGCTGTGTGACTGCTGCGAGAGGATTTGTA 268
QY 169 CAGTATTTGCGCTGCAATGCAAAAAAGCTGCTGCATGCAAGAGCAAAAGTATTTGA 228
Db 269 AACCTATGCTCTTCTTCCTTGCGCAAAAGCTGCGCAAGGCTTGTGTAACCTATGTAAGT 328

QY 229 TTGACCGGGAAGGCGAACAACTGCACAAGATTGTCGTGAGTTAGCGAAGACCGTACG 288
Db 329 CAGCAAGAGAGGCTGAGGAAGTTTCTAAACAGATTGAAGCATATGAGGCCAGGCTATTA 388
QY 289 CGCTGCACTCGATCTCTTAATATCAGCAAGTCCATTAACATGCTGGCGCAATATATG 348
Db 389 CTTTGGGGGTGATGCTCCCAAGAGGCTGATGTGAAGCCATGATGAAAAAGCGCTATTG 448
QY 349 AACTGGCGGTGGCTGATATTTTTCATGCAATGCAAGCGGCTTATTTGCGGCCAG 408
Db 449 ATGATGAGGGAACCATTTGATGCTGCTCAACATGCAAGCAATCCTCGGATACCTTGT 508
QY 409 TGCGTGAAGGTGATCCAGATGCTGCGTGTAAATTCATTAATATATTCGCGGT 468
Db 509 TGATACGAATGAAGAAGTCCCAATGGATGAAGTATTTGAATTCACCTGAGACTGT 568
QY 469 TTGCGTGTGTCCTGCAAGTCTGCGCATATGATTTGGCGAGAGTGGCGGATATATTT 528
Db 569 TTCTGTGTACCCAGGCGCAACAAAGTCATGTGAAGAAGCAAGGAAGATCATCA 628
QY 529 TTACCAAGTTCATCGCGCGGCTGCTCCGCTTATTTGGAACCGATCTACCGCGTCCA 588
Db 629 ACATTTGCGTACGTTGTTGCTCATTTGTAATTTGGCCAAAGCAATACGCTGCTGTA 688
QY 589 AATTGGCGTTCAAGCATTCGTACACACTACCGCGCGCAGGTTTCTCAATATGCGTGC 648
Db 689 AACCTGCTTATTTGGGTTCTCCAAAGACTGCCCGCAGAGAGGCTGCGAGCATATTA 748
QY 649 GTGTGGGTGGGCTGCTGCCAGACCACTACTGCTGCTGTAATGACTGGCCAAAG 708
Db 749 ATGTCATGTTGTTGGCTTCAATGATGATGATGATGATGATGATGATGATGATGAT 808
QY 709 CCAAAATGGAAGAGCCCTGCAAA 732
Db 809 ACATGGAAGAAATCTTGGAA 832

RESULT 9
US-08-793-035-1
; Sequence 1, Application US/08793035
; Patent No. 6011201
; GENERAL INFORMATION:
; APPLICANT: Slabas, Antoni R.
; APPLICANT: White, Andrew
; APPLICANT: Chase, Dianne
; APPLICANT: Elborough, Keiran
; APPLICANT: Fentem, Phillip A.
; TITLE OF INVENTION: B-ketocyl ACP Reductase Genes from
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,035
; FILING DATE: 28-JUL-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9414622.2
; FILING DATE: 20-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB95/01678
; FILING DATE: 17-JUL-1995

ATTORNEY/AGENT INFORMATION:
 NAME: Kammerer, Patricia A.
 REGISTRATION NUMBER: 29,775
 REFERENCE/DOCKET NUMBER: MOBT:132
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 713.787.1440
 TELEFAX: 713.787.1440

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:

LENGTH: 1185 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-793-035-1

Query Match 1.4%; Score 56; DB 3; Length 1185;
 Best Local Similarity 43.1%; Pred. No. 1.9e-07;
 Matches 269; Conservative 0; Mismatches 355; Indels 0; Gaps 0;

109 CTGTGCCCCCTATGAACTACCCCTTAATGCAAGTTGACGATACCTGCGCTGCGT 168
 266 CTCTTCCGAAGTGAAGTCCCGGTGCTGTGTGACTGCTTGCAGAGGATTTGTA 325
 169 CAGGATTTGGCTGCAATGTGCAAAAACGCTGCTGATGAGAGCAAGAGTATTTGA 228
 326 AACCTATTGCTCTTCTCTTGCGCAAGAGCTGCAAGGCTTGTGTAATGCTAGGT 385
 229 TTGACCGGGAGAGCGCAAACTGACAAAGTTTCCCTGAGTTAGCGCAAAAGCGGTACG 288
 386 CAGCAAAAGAGGCTGAGAGGATTTCTAAGATTTAAGCATTTGAGGCGCAGCTATTGA 445
 289 CCGTGCACCTGATCTCTTCAATATACAGAACTGATACATGCGGCGGACATTATCG 348
 446 CTTTGGGGGTGATGTCTCCAAAGAGCTGATGTGGAAGCATGATGAAACCGCTATTG 505
 349 AACTGCGGGGTGGCTGATATTTTTCATGCAATGCGAGGCGCTTATTTGGCGGCGAG 408
 506 ATGCATGAGGAGAACCATTTGATGCTGCTCAACATGCGAGGAAATCAGTCGGGATACCTTGT 565
 409 TGGCTAAGGTGATCCAGATGTCTGGATGCTGTTAATCTGAATATATAATGCGGCGT 468
 566 TGATACGAATGAAGAAGTCCCAATGGGATGAAGTATGTTGTAATCTCACTGAGTCT 625
 469 TTGCTGTGCTCGTCAAGTCTCCGCAATGATTTGGCGAGAGTGGCGGATATATTT 528
 626 TTCTCTGACCCAGGAGCAACAAGATCTGATGAAGAAGAGAGATCATCTCA 685
 529 TTAACGATTCATCGCGGGGCTGCTTCCGTTATCTGGGAACCGATCTACCGCGCTCA 588
 686 ACATTCGCTGATGTTGTTGCTCATGTTGATTAATTTGGCCAAAGCAACTACGCTGCTCA 745
 589 AATTTGCCCTTCAGGCAATTTGTACACATCCCGCGCCGAGTTTCTCATATGCGCTGC 648
 746 AACCTGCTGTTATTTGTTCTTCCAAAGCTGCGCCGAGAGGAGTGCAGAGAGATATAA 805
 649 GTGTGGGTGGGTGCTGCTCCAGACAGTACATCACTGCTGCTTGTGATGATGCGCAAAAG 708
 806 ATGTCAATGCTGTTGCTGCGGCTGCTCATTTGATCATCTGACATGACTGCCAACCTTGGAGAAG 865
 709 CCAAAATGGAAGAGCCCTGGCAA 732
 866 ACATGGAAGAAATCTTGGAA 889

RESULT 10
 US-08-594-808B-6
 Sequence 6, Application US/08594808B
 Patent No. 5804423
 GENERAL INFORMATION:
 APPLICANT: Klasen, Ralf
 APPLICANT: Bringer-Meyer, Stephanie
 APPLICANT: Sahm, Hermann
 APPLICANT: Hollenberg, Cornelies P

TITLE OF INVENTION: MICROBIOLOGICAL METHOD OF MAKING
 NUMBER OF INVENTION: 5-KETOGLUCONATE
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: The Firm of Karl F. Ross, PC
 STREET: 5676 Riverdale Ave.
 CITY: Bronx
 STATE: New York
 COUNTRY: USA
 ZIP: 10471-0900

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/594,808B

FILING DATE: 07-FEB-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Myers, Jonathan E

REGISTRATION NUMBER: 26,963

REFERENCE/DOCKET NUMBER: 19893

TELECOMMUNICATION INFORMATION:

TELEPHONE: (718) 884-6600

TELEFAX: 718/601-1099

TELEX: 620428

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1512 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 280..1050

US-08-594-808B-6

Query Match 1.4%; Score 53.8; DB 1; Length 1512;
 Best Local Similarity 46.6%; Pred. No. 1.2e-06;
 Matches 215; Conservative 0; Mismatches 237; Indels 9; Gaps 1;

147 GCAGCTATCTGCGCGTCAAGTATGCGCTGCAATGTGCAAAAGCTGCTCGAT 206
 319 GCCCTGTGACCGGCGGTACGCGGGAATTTGGCTGACGCTGGCAAGGGGCTTGGCGGC 378
 207 GCAGGACCAAGGTAGTATTTGATTTGACCGGGAAGCGCAAACT-----GCACAG 257
 379 TACGGGCTAGAGTTGCTGTAATGCGGGAATGCTGAAGTGTGACTGCGCAGTCC 438
 258 ATTGTGCTAGTTAGCGCAAAAGCGCTACGCGCTGCACTGATCTTCAATATACAG 317
 439 GGGTTGAGCGGGAAGATTGAAGCCAGTACGGCGGTTTTCATGTGACGATACAGAT 498
 318 CAACTGATTAACATGCTGGGCGGACATTATCGAACTGGCGGGGTGGCTGGATTTTTCAT 377
 499 GCGGTATTTATGCGCGTGGCGGATTTGAGCGGGAACATTTGGACGATGATTCCTGATC 558
 378 GCCAATGACGAGCGCTTATTTGGCGGCCAGTGGCTGAAGTGTACAGATGCTGGGAT 437
 559 AATTAATCCGGATACAGCGCCGAGCGCTCTGTGAGAGTTTTCGCGCAAGACTGGGAT 618
 438 CGGTGTTAAATGTGAATTAATGCGCGCTTTCGCTGTGCTGCGTCACTGCGCGCAT 497
 619 GATCTGATGTCACCAATGTCAACGCGGTTTCTGTCGCGGAGCGGCGTGGCGGCGCAC 678
 498 ATGATTTGCGAGAGTGGCGGATATATTTTACAGTTTCCATCGCGGGCGTCTGTTCCG 557
 679 ATGATTTCCCGGAGCGGCGCAAGATGTCAAATATCTGTTCCGTCAGAGTGAATCGCC 738

Dy 558 GTTATCTGGGAAACCATCTTACACCCCGTCCAATTGGCGGT 598
Db 739 GGTCCGGGAATTGCCTTATACGGCGACCAAGGAGCGGT 779

RESULT 11
US-09-026-482B-1
; Sequence 1, Application US/09026482B

Query Match	1.3%	Score 52.2	DB 3	Length 1670
Best Local Similarity	47.5%	Pred. No. 4.1e-06		
Matches 233, Conservative	0	Mismatches 243	Indels 15	Gaps 2

Accession	Sequence	Position
Qy	TCCGCGCAGTCTCTCGCGGATATGATTTGGGCGAGAGTCGGGGGATATATATTTTTCACAGTT	533
478		11111
Db	TGTTAATTTTCTGCGCACAATATGATTAAGCTTAAATTAATGGCGACATCATCAATATACAGCT	81.4
755		11111
Qy	CCATTCGGCGGCGGCTGTTCCGGTTATCTGGGAAACGATCTACCGCGTCCCAATTTGCGG	597
538		11111
Db	CTATTGGGTATTTGGCCCAATGCGACCCCTTTTCTGCTATATGTCGCTTAAAGCTTCGC	87.4
815		11111

QY	598	TTCAGGCATTTC	608
Db	875	TGCATGCCCTTC	885

RESULT 12
US-09-296-284-67
; Sequence 67, Application US/09296284A

Query Match	1.3%	Score 51.8;	DB 4;	Length 634;
Best Local Similarly	50.68;	Pred. No. 2.8e-06;		
Matches 125; Conservative	0;	Mismatches 122;	Indels 0;	Gaps 0

OY	523	TAAATTTTACCAAGTTTCATCTCGGGGCGTGCTCCGGTTACTGTGGAAACCATTACACCG	582
Dd	192	TCAATCACAATCATAGTTTCCAGGCCACAGCTCTGTGCGCGATTTAGACCAATGATGCTTAACCTCG	251
OY	583	CGTCCAA	589
Dd	252	CATCOAA	258

RESULT 13
 US-09-296-284-66
 Sequence 66, Application US/09296284A
 Patent No. 6204040
 GENERAL INFORMATION:
 APPLICANT: Choi, Eun-Sung
 APPLICANT: Rhee, Sang-Ki
 APPLICANT: Lee, Eun-Hae
 TITLE OF INVENTION: Glucuronobacter Suboxydans Sorbitol Dehydrogenase, Genes
 TITLE OF INVENTION: and Methods of Use Thereof
 FILE REFERENCE: 1533.0870000
 CURRENT APPLICATION NUMBER: US/09/296,284A
 CURRENT FILING DATE: 1999-04-22
 NUMBER OF SEQ ID NOS: 87

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 11:05:17 ; Search time 296.853 Seconds
(without alignments)
10373.549 Million cell updates/sec

Title: US-09-802-208B-2
Perfect score: 3971
Sequence: 1 atcgattgagagtgcttctt.....tcacgcgcgtcacatcgat 3971

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCIT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCITUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3971	100.0	3971	US-09-802-208B-2	Sequence 2, Appli
2	101	2.5	417	US-09-960-352-11317	Sequence 11317, A
3	71.2	1.8	1335	US-09-738-626-127	Sequence 127, App
4	67	1.7	6133	US-10-114-1170-15	Sequence 15, Appl
5	64.6	1.6	789	US-09-815-242-6500	Sequence 6500, Ap
6	64.6	1.6	29729	US-09-070-927A-238	Sequence 238, App
7	63.4	1.6	468	US-09-974-300-2287	Sequence 2287, Ap
8	63.2	1.6	744	US-09-974-300-6558	Sequence 6558, Ap
9	58	1.5	807	US-09-712-363-99	Sequence 99, Appl
10	56.4	1.4	8320	US-09-813-453A-71	Sequence 71, Appl
11	55.6	1.4	6948	US-09-070-927A-4	Sequence 4, Appli
12	54.4	1.4	537	US-09-974-300-6537	Sequence 6537, Ap
13	50.8	1.3	1872	US-09-885-303A-15	Sequence 15, Appl
14	50.6	1.3	1714	US-09-885-303A-13	Sequence 13, Appl
15	50.2	1.3	680	US-09-974-300-2280	Sequence 2280, Ap
16	49.4	1.2	11427	US-09-070-927A-165	Sequence 165, App
17	48.8	1.2	3153	US-09-922-501-13	Sequence 13, Appl
18	48.2	1.2	735	US-09-815-242-6029	Sequence 6029, Ap
19	47.2	1.2	744	US-09-815-242-7795	Sequence 7795, Ap

20	45.8	1.2	756	10	US-09-974-300-2400	Sequence 2400, Ap
21	45	1.1	765	9	US-09-978-758-1	Sequence 1, Appli
22	44.4	1.1	774	9	US-09-891-6641-55	Sequence 55, Appl
23	44.4	1.1	1043	10	US-09-823-901-1	Sequence 1, Appli
24	43.4	1.1	741	10	US-09-815-242-8705	Sequence 8705, Ap
25	43.4	1.1	741	10	US-09-815-242-9002	Sequence 9002, Ap
26	42.4	1.1	31728	9	US-10-114-170-64	Sequence 64, Appl
27	42.2	1.1	672	10	US-09-974-300-6548	Sequence 6548, Ap
28	41.6	1.0	765	10	US-09-974-300-2256	Sequence 2256, Ap
29	41.4	1.0	14286	10	US-09-070-927A-162	Sequence 162, App
30	41	1.0	732	10	US-09-815-242-9262	Sequence 9262, Ap
31	41	1.0	713	10	US-09-815-242-9483	Sequence 9483, Ap
32	40.8	1.0	714	10	US-09-823-901-3	Sequence 3, Appli
33	40.8	1.0	717	10	US-09-974-300-2252	Sequence 2252, Ap
34	40.6	1.0	418	10	US-09-880-107-529	Sequence 529, App
35	40.4	1.0	708	9	US-09-738-626-2684	Sequence 2684, Ap
36	40.4	1.0	833	10	US-09-770-445-686	Sequence 686, App
37	40.2	1.0	599	9	US-09-978-758-11	Sequence 11, Appl
38	40.2	1.0	738	10	US-09-815-242-4406	Sequence 4406, Ap
39	40.2	1.0	741	10	US-09-815-242-8025	Sequence 8025, Ap
40	40.2	1.0	1905	10	US-09-922-488-3	Sequence 3, Appli
41	40.2	1.0	807	9	US-09-738-626-1473	Sequence 1473, Ap
42	40	1.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
43	39.8	1.0	762	10	US-09-974-300-2259	Sequence 2259, Ap
44	39.8	1.0	816	10	US-09-815-242-9289	Sequence 9289, Ap
45	39.8	1.0	816	10	US-09-815-242-9581	Sequence 9581, Ap

ALIGNMENTS

RESULT 1
US-09-802-208B-2
Sequence 2, Application US/09802208B
Publication No. US20030041352A1

GENERAL INFORMATION:
APPLICANT: Pariotti, Wayne
APPLICANT: Lafayette, Peter
APPLICANT: Kane, Patrick
TITLE OF INVENTION: Arbidol or Ribitol As Positive Selectable Markers
FILE REFERENCE: UGA 85SR
CURRENT APPLICATION NUMBER: US/09/802, 208B
CURRENT FILING DATE: 2001-03-08
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0

SEQ ID NO 2
LENGTH: 3971
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: misc_feature
LOCATION: (96)..(848)
OTHER INFORMATION: ribitol dehydrogenase coding region

FEATURE:
NAME/KEY: misc_feature
LOCATION: (859)..(2463)
OTHER INFORMATION: ribitol kinase coding region

NAME/KEY: misc_feature
LOCATION: (2565)..(3839)
OTHER INFORMATION: ribitol transporter coding region

US-09-802-208B-2

Query Match 100.0%, Score 3971, DB 9, Length 3971:
Best Local Similarity 100.0%, Pred. No. 0;
Matches 3971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCGATTGACGAGTTGCTTCACACGCGATTAATTCGCTCAGTGCCTGCAACAA 60
|||||
DB 1 ATCGATTGACGAGTTGCTTCACACGCGATTAATTCGCTCAGTGCCTGCAACAA 60
|||||

OY 61 CAGAGCTAATAATTGCGCCCTGTAAAGATTAATGATGATCACTGTGCCCTCTA 120
|||||

Db 61 CAGAGACTAATAATTCGCCCTGGTAAAGATTATATGATGAATCACTGTGTCCTCTA 120
Qy 121 TGAATACCTCCCTTAAATGCGAAAGTTGACAGCTATCACTGGCGTGCCTGAGTATTGGC 180
Db 121 TGAATACCTCCCTTAAATGCGAAAGTTGACAGCTATCACTGGCGTGCCTGAGTATTGGC 180
Qy 181 TGCATATGCGAAAGAGCTGCTGATGACAGAGCAAGGATGATTTGATTGATTTGACCGGAG 240
Db 181 TGCATATGCGAAAGAGCTGCTGATGACAGAGCAAGGATGATTTGATTGATTTGACCGGAG 240
Qy 241 GCGACAACAGCGCAAGATTTGCTGATGATTTAGCGGAAAGCGGTACGGCTGCAACTGC 300
Db 241 GCGACAACAGCGCAAGATTTGCTGATGATTTAGCGGAAAGCGGTACGGCTGCAACTGC 300
Qy 301 ATCTCTTCAATTAATCAGCAAGTGCATTAACATGCTGGCGGCAATTAATCAGTGCAGG 360
Db 301 ATCTCTTCAATTAATCAGCAAGTGCATTAACATGCTGGCGGCAATTAATCAGTGCAGG 360
Qy 361 GGGTGGATATTTTTCATGCGCAATGCGAGGCGCTTATATTGGCGGCGCAATGCTGAGG 420
Db 361 GGGTGGATATTTTTCATGCGCAATGCGAGGCGCTTATATTGGCGGCGCAATGCTGAGG 420
Qy 421 ATCCAGATGCTGAGATCGTGTGTTAAATCTGAATATAAATGGCGGCTTCGGTGTGTC 480
Db 421 ATCCAGATGCTGAGATCGTGTGTTAAATCTGAATATAAATGGCGGCTTCGGTGTGTC 480
Qy 481 GTGCAAGTCTGCGGCATATGATGCGAGAGTGGCGGCAATATATTTTACAGATTCCA 540
Db 481 GTGCAAGTCTGCGGCATATGATGCGAGAGTGGCGGCAATATATTTTACAGATTCCA 540
Qy 541 TGCAGGCGCGTGTCCGGTATCTGGGAAACCGATCTACACCGGCTTCCAAATTTGCCCTTC 600
Db 541 TGCAGGCGCGTGTCCGGTATCTGGGAAACCGATCTACACCGGCTTCCAAATTTGCCCTTC 600
Qy 601 AGGCAATTCGTAACACATACACCGCGCGCAGGTTTCTCAATATGGCGTGCCTGAGGTCGG 660
Db 601 AGGCAATTCGTAACACATACACCGCGCGCAGGTTTCTCAATATGGCGTGCCTGAGGTCGG 660
Qy 661 TGCCTGCGAGGACAGTATGATGCTGCTGCTGATGATGCTGCGCAAAAGCCAAATGAGAG 720
Db 661 TGCCTGCGAGGACAGTATGATGCTGCTGCTGATGATGCTGCGCAAAAGCCAAATGAGAG 720
Qy 721 AAGCCCTGGCAAAATGATGAGCTGATGCAACCGATTTGAAGTGGGGAATCGATTTGTTTA 780
Db 721 AAGCCCTGGCAAAATGATGAGCTGATGCAACCGATTTGAAGTGGGGAATCGATTTGTTTA 780
Qy 781 TGTGACCCGCTGAAAAATGTCACCGTGCAGATTTAGTATGATCTGCTGCGCAGTGTGC 840
Db 781 TGTGACCCGCTGAAAAATGTCACCGTGCAGATTTAGTATGATCTGCTGCGCAGTGTGC 840
Qy 841 ATCTGTAAGGCGCATATGACAAATACCAAAACGTTTGGTGTGATTTGGGATCA 900
Db 841 ATCTGTAAGGCGCATATGACAAATACCAAAACGTTTGGTGTGATTTGGGATCA 900
Qy 901 GGCAGTGTCCGCGCGGAGATTTTGTATCTCAACGATCTGTGCTATCCATGCCACAGAA 960
Db 901 GGCAGTGTCCGCGCGGAGATTTTGTATCTCAACGATCTGTGCTATCCATGCCACAGAA 960
Qy 961 AAAATCAGCATACGCGCGCGCAGCGAAAGCGGTGGAACAGTCCAGCCAGAGATGTGG 1020
Db 961 AAAATCAGCATACGCGCGCGCAGCGAAAGCGGTGGAACAGTCCAGCCAGAGATGTGG 1020
Qy 1021 CAGGCGGTCTGTTCATGATTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 1021 CAGGCGGTCTGTTCATGATTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Qy 1081 GTGGCAGGCAATGCTGTTTGTATGCGACCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1081 GTGGCAGGCAATGCTGTTTGTATGCGACCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Qy 1141 CCATTCCTGTTCAGCCCGGAGAGATGCAAAACCAATATCATTTGTGTGATGATGATCAC 1200
Db 1141 CCATTCCTGTTCAGCCCGGAGAGATGCAAAACCAATATCATTTGTGTGATGATGATCAC 1200

Qy 1201 CCGCGCACCGAACAGCAGAGCAATCAATGCCACTCCACATCCGGTGTGAATACGTC 1260
Db 1201 CCGCGCACCGAACAGCAGAGCAATCAATGCCACTCCACATCCGGTGTGAATACGTC 1260
Qy 1261 GGTGTGTAATTTTCCCTGTAATGGAACACCGAAATTTCTGTGGCTGGAAGAAATATG 1320
Db 1261 GGTGTGTAATTTTCCCTGTAATGGAACACCGAAATTTCTGTGGCTGGAAGAAATATG 1320
Qy 1321 CCAGAGATCTACGAAAGCTGCGGCAATTTTGTGATCTGCGCGATTTTGTGACCTGGCG 1380
Db 1321 CCAGAGATCTACGAAAGCTGCGGCAATTTTGTGATCTGCGCGATTTTGTGACCTGGCG 1380
Qy 1381 GCTAACCGGTATTTAGCGCTTCAGTATGCACTGTTAACTGTAATGAGAGCTGGTGCA 1440
Db 1381 GCTAACCGGTATTTAGCGCTTCAGTATGCACTGTTAACTGTAATGAGAGCTGGTGCA 1440
Qy 1441 CATGAAATATGCTGGGATTCAGATTTATTTCCGACCAATTTGGCTGTGAGATGAGCGAT 1500
Db 1441 CATGAAATATGCTGGGATTCAGATTTATTTCCGACCAATTTGGCTGTGAGATGAGCGAT 1500
Qy 1501 GAAGATTTTATTCGATTTGCTATCATATGTTTCTCCGGAACACCTTGTGAAATGAT 1560
Db 1501 GAAGATTTTATTCGATTTGCTATCATATGTTTCTCCGGAACACCTTGTGAAATGAT 1560
Qy 1561 TTAACAGCACAAGCGCGCGAGAGATGGAATTAATCCCGGCAACCTTGTGCTGTAAGG 1620
Db 1561 TTAACAGCACAAGCGCGCGAGAGATGGAATTAATCCCGGCAACCTTGTGCTGTAAGG 1620
Qy 1621 TTAATGATGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 1621 TTAATGATGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Qy 1681 AATCTCGCTATGTTTTCGCGCACTTCTTCAATGCAACATGCAATCTACACTTCTCCCTCG 1740
Db 1681 AATCTCGCTATGTTTTCGCGCACTTCTTCAATGCAACATGCAATCTACACTTCTCCCTCG 1740
Qy 1741 TTTGTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Db 1741 TTTGTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Qy 1801 GAAGCGGCGCAAAAGTGTGCGGAGAGCTATTGACAGCTACTGATTTTCAATCCGCT 1860
Db 1801 GAAGCGGCGCAAAAGTGTGCGGAGAGCTATTGACAGCTACTGATTTTCAATCCGCT 1860
Qy 1861 GTTGAAGAGCTGCGCAATATGCAAAAGTGTGAATCAGCCCTTCCCTGCTGCTGCTGCT 1920
Db 1861 GTTGAAGAGCTGCGCAATATGCAAAAGTGTGAATCAGCCCTTCCCTGCTGCTGCTGCT 1920
Qy 1921 GATCGAATCTCGCAAAAGAGGCGCAACCATGATGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db 1921 GATCGAATCTCGCAAAAGAGGCGCAACCATGATGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Qy 1981 CAGTGTGCTGCGCAATTTCTGGAATTCGCGCCCTTCGAGATCTCTCATATGCGCAGAG 2040
Db 1981 CAGTGTGCTGCGCAATTTCTGGAATTCGCGCCCTTCGAGATCTCTCATATGCGCAGAG 2040
Qy 2041 GATATTTGCTGCGGATATGAGAGCGAGATCTGGAATTAATTTACTGCTGCTGCTGCTGCT 2100
Db 2041 GATATTTGCTGCGGATATGAGAGCGAGATCTGGAATTAATTTACTGCTGCTGCTGCTGCT 2100
Qy 2101 GGAATTTAGCGGAAATTTGCTGCTGCGCAAAATTTCTCAGCCTCTCAAAAGCGAGGGA 2160
Db 2101 GGAATTTAGCGGAAATTTGCTGCTGCGCAAAATTTCTCAGCCTCTCAAAAGCGAGGGA 2160
Qy 2161 GTAGTGTGTAATAATATCTTATTTAGCGGCGGAGAGATCCATGCTGATACAGCA 2220
Db 2161 GTAGTGTGTAATAATATCTTATTTAGCGGCGGAGAGATCCATGCTGATACAGCA 2220
Qy 2221 ATTCTGCGAGATACCTGCGGATTTCCGCTATTTACAGCAAGATGCTGCGAAGCGGTTT 2280
Db 2221 ATTCTGCGAGATACCTGCGGATTTCCGCTATTTACAGCAAGATGCTGCGAAGCGGTTT 2280

Qy 2281 TTAGGCTGGCCATTTCTGCTGTGCGCGGAATATTGCACCTCTGTGTGGCGAAGCG 2240
|||||
Db 2281 TTAGGCTGGCCATTTCTGCTGTGCGCGGAATATTGCACCTCTGTGTGGCGAAGCG 2240
Qy 2341 ATGCACCAATTTACCCATGTGATTAATATTATATCCGCAAGAACGCTATCATCTCTT 2400
|||||
Db 2341 ATGCACCAATTTACCCATGTGATTAATATTATATCCGCAAGAACGCTATCATCTCTT 2400
Qy 2401 CATCATCGTCGATATGAGGCGCTTAAGCAAGTTGACAGCTACATCTCAAAATTTACTCAGAGAC 2460
|||||
Db 2401 CATCATCGTCGATATGAGGCGCTTAAGCAAGTTGACAGCTACATCTCAAAATTTACTCAGAGAC 2460
Qy 2461 TAATTAACAGCGCGCTGAAGCTGTTTTCAGGCAATCACTAATTAAGACTCACTCCGCTA 2520
|||||
Db 2461 TAATTAACAGCGCGCTGAAGCTGTTTTCAGGCAATCACTAATTAAGACTCACTCCGCTA 2520
Qy 2521 ATATCCCGAGAGTCATTCATCTGACCCCTTAAGCAAGTTTATATGTCAGAAATTAATA 2580
|||||
Db 2521 ATATCCCGAGAGTCATTCATCTGACCCCTTAAGCAAGTTTATATGTCAGAAATTAATA 2580
Qy 2581 AACAGTGGTTGGGTTGGCACTGATCTGATATGGGATATATCGGCATCGCATTTTAA 2640
|||||
Db 2581 AACAGTGGTTGGGTTGGCACTGATCTGATATGGGATATATCGGCATCGCATTTTAA 2640
Qy 2641 TGAATGATGATGTTTTCGAACCTGATCTTATCGCATATATTAATCGTTAGGCTTCA 2700
|||||
Db 2641 TGAATGATGATGTTTTCGAACCTGATCTTATCGCATATATTAATCGTTAGGCTTCA 2700
Qy 2701 CACCTGCGGAAGCCTCTTTTGGCTTTACGCTCTACAGGCGCTGGCGCTTCCGCTT 2760
|||||
Db 2701 CACCTGCGGAAGCCTCTTTTGGCTTTACGCTCTACAGGCGCTGGCGCTTCCGCTT 2760
Qy 2761 GGGTTCTGGGGATGAGGGGAATCATCAGCGCGCAAAAACATGCTGATTTGGTTTG 2820
|||||
Db 2761 GGGTTCTGGGGATGAGGGGAATCATCAGCGCGCAAAAACATGCTGATTTGGTTTG 2820
Qy 2821 TCCATATGATGCGATTTTCATGTTCTGTTCTGTTGATGAGGACAGCAAGCAATATG 2880
|||||
Db 2821 TCCATATGATGCGATTTTCATGTTCTGTTCTGTTGATGAGGACAGCAAGCAATATG 2880
Qy 2881 GGTATATCTGCTGTTTACGGGATTCGTTGTTAGCGTATCCGCTATTTCTTACTCAT 2940
|||||
Db 2881 GGTATATCTGCTGTTTACGGGATTCGTTGTTAGCGTATCCGCTATTTCTTACTCAT 2940
Qy 2941 TTATCGTCTTATTTATTCATTAAGCTGATGAGCAAAATTCAGTTGCTGGTTGGT 3000
|||||
Db 2941 TTATCGTCTTATTTATTCATTAAGCTGATGAGCAAAATTCAGTTGCTGGTTGGT 3000
Qy 3001 ACTGGCGGATATATTCAGTATGAGGATGCTGCTGCGACATTAATTCAGATTTTACGA 3060
|||||
Db 3001 ACTGGCGGATATATTCAGTATGAGGATGCTGCTGCGACATTAATTCAGATTTTACGA 3060
Qy 3061 TACCGATTAATGCTGAATATGGAACCTTATGCTGACACTGCGCTTCTGCTGGCG 3120
|||||
Db 3061 TACCGATTAATGCTGAATATGGAACCTTATGCTGACACTGCGCTTCTGCTGGCG 3120
Qy 3121 GGTGATGTCATGATCTCTTCCGTCATGTTAAACGCGCTGACATATGCAATATTTAA 3180
|||||
Db 3121 GGTGATGTCATGATCTCTTCCGTCATGTTAAACGCGCTGACATATGCAATATTTAA 3180
Qy 3181 CTCGCCGTAGAGATTTGAGATTAAGTCGGGACAGTAATTTACTTTATACCAACGCA 3240
|||||
Db 3181 CTCGCCGTAGAGATTTGAGATTAAGTCGGGACAGTAATTTACTTTATACCAACGCA 3240
Qy 3241 ATATTTTCTCTCCAGTATTTGCGCATTAATAATACCTTATATTCGTTTGGCG 3300
|||||
Db 3241 ATATTTTCTCTCCAGTATTTGCGCATTAATAATACCTTATATTCGTTTGGCG 3300
Qy 3301 TCAATTAAGCAATGATTTGATGAGCACTGGATTCACCACTCGAATGTTGAGG 3360
|||||
Db 3301 TCAATTAAGCAATGATTTGATGAGCACTGGATTCACCACTCGAATGTTGAGG 3360
Qy 3361 TCTGGCGGCAATTTTCTTACCACTATTTTCTAATATTTTGGGGGATGTCGAC 3420

Db 3361 TCTGGCGGCAATTTTCTTACCACTATTTTCTAATATTTTGGGGGATGTCGAC 3420
Qy 3421 AAAAATGGGCTGATGCGGTTATTCGCTGGTTGGTCCCGGAGTGGCAATCA 3480
|||||
Db 3421 AAAAATGGGCTGATGCGGTTATTCGCTGGTTGGTCCCGGAGTGGCAATCA 3480
Qy 3481 GTTTAGCGTTTCTACATGCGGCAATTTGTTGTCACAACTCTGATGAGCAATATTC 3540
|||||
Db 3481 GTTTAGCGTTTCTACATGCGGCAATTTGTTGTTGTCACAACTCTGATGAGCAATATTC 3540
Qy 3541 CGGCAATGCTCTGGAACCTTTTGTGTCGATTTGTCCGATGCGCGCTTCTCCCG 3600
|||||
Db 3541 CGGCAATGCTCTGGAACCTTTTGTGTCGATTTGTCCGATGCGCGCTTCTCCCG 3600
Qy 3601 CACTGAAACCAAAACAAAGGTCGCAATCTCGTTTACACCTCTGCGGGATATG 3660
|||||
Db 3601 CACTGAAACCAAAACAAAGGTCGCAATCTCGTTTACACCTCTGCGGGATATG 3660
Qy 3661 CTAACTTCTGCTCCGCGCAATTTGCGGTGTTATTAACCGTGTAGCATATCGTG 3720
|||||
Db 3661 CTAACTTCTGCTCCGCGCAATTTGCGGTGTTATTAACCGTGTAGCATATCGTG 3720
Qy 3721 TGGTCAATGCTTATACAGCATTTGATCTATTTGGCTTTGCTTGGCATTCATTCGCG 3780
|||||
Db 3721 TGGTCAATGCTTATACAGCATTTGATCTATTTGGCTTTGCTTGGCATTCATTCGCG 3780
Qy 3781 TTGAGCAGCAGATTCAGTTCGCGCAATGCTGAGAGGATTAATCTCTGAA 3840
|||||
Db 3781 TTGAGCAGCAGATTCAGTTCGCGCAATGCTGAGAGGATTAATCTCTGAA 3840
Qy 3841 AAACGAAGCAGATTCAGTTCGCTCTCTCATGAGGAGATTAATTCATCAGG 3900
|||||
Db 3841 AAACGAAGCAGATTCAGTTCGCTCTCTCATGAGGAGATTAATTCATCAGG 3900
Qy 3901 CAATAGTACTTGTATGAGATTAACGTCGACGCGGTTAATCAGTTTACGCGCT 3960
|||||
Db 3901 CAATAGTACTTGTATGAGATTAACGTCGACGCGGTTAATCAGTTTACGCGCT 3960
Qy 3961 CAGCCATCGAT 3971
|||||
Db 3961 CAGCCATCGAT 3971

RESULT 2
US-09-960-352-11317
; Sequence 11317, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11317
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB34-080-Q1-E1-D8
US-09-960-352-11317

Query Match 12.5%; Score 101; DB 10; Length 417;
Best Local Similarity 56.2%; Pred. No. 2, 7e-22;
Matches 232; Conservative 0; Mismatches 175; Indels 6; Gaps 2;
Qy 1194 GGATCACCGCGCGTCAAGTCAGTCACCGAGATCAACGAGACACAGCAGCTCTCGCA 61
Db 2 GGATCACCGCGCGTCAAGTCAGTCACCGAGATCAACGAGACACAGCAGCTCTCGCA 61

RESULT 4

US-10-114-170-15

Sequence 15, Application US/10114170
Publication No. US20030023075A1

GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESSES:

ADDRESS: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296,95017TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 6133
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linearMOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-114-170-15Query Match 1.7%; Score 67; DB 9; Length 6133;
Best Local Similarity 47.7%; Pred. No. 3.4e-10;
Matches 236; Conservative 0; Mismatches 250; Indels 9; Gaps 1;

QY 119 TATGAATACCTCCCTTAATGCAAGTTCACATGCTAGTGGCGGCTCAGGATTTGG 178
DB 644 TTGTAAGGAGCAAGTAATGATAGGTCAATTAATACCGCGCATCAAGTGTATGG 703
QY 179 CCTGCAATGTGCAAAACCGTCTGATGACAGAGCAAGAGTATGATGACCGGA 238
DB 704 GGAAGGATATGCCAGAGCTTGAATGACAGTGCAGAGTTTACGGAACGAG 763
QY 239 AGCGCAACATGCAACAGATTTGCTGAGTT-----AGCGGAACCGGTACGC 289
DB 754 AGTTAGGCTATCGAAGCATTCGCAACCAATCTGCCCGGAGGAGGATTTGCTAAAGC 823
QY 290 GGTGCACTGCATCTTCAATATACGACAGATGATATACATGCTGGCGGACATTAATGA 349
DB 824 GCGGAGATTGGATGTACACAGACGACAGTTCATGGCCGATTTGTCGAAGCAGCGTTGA 883
QY 350 ACTGGCGGTGGCTGATATTTTCATGCAATGACAGCGCTTATATTGGCGGCCACT 409

DB 884 TAGCTGGGGCGAGTTGATGTTCTTATCAATATGCGGGCGTTATGCCCTTTCACCCGT 943
QY 410 GGCTAGAGTGAATCCAGATGCTGGATGCTGTGTTAAATCGAATATAATAGCGCGCT 469
DB 944 TGCAGAGGAAAAACAGAGTAATGCGGCTCACTATGACGTAATATCAAGGTGACT 1003
QY 470 TCGCTGTGCGGTGACAGTCCGCGCATATGATGATGGCGAGAGTGGCGGATATAATTT 529
DB 1004 GTGGGGAATTTGGCGGTCTTCTCCGGTATGACAGCAGAGGTTCCGGTCAATATCA 1063
QY 530 TACCAATTCATCGGGGCGGTGTTCCGGTATCTGGGAACGATCTACACCGGTCCAA 589
DB 1064 TCTGTGTTATTTGGTGGCCCTTCTGTGTGTGCCACAGCGCAGTCTATTTGCAATCAA 1123
QY 590 ATTGCGGTTACGC 604
DB 1124 ATTGCGAGTACGGCG 1138

RESULT 5

US-09-815-242-6500

Sequence 6500, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6500

LENGTH: 789

TYPE: DNA

ORGANISM: Enterococcus faecalis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(789)

US-09-815-242-6500

Query Match 1.6%; Score 64.6; DB 10; Length 789;
Best Local Similarity 48.7%; Pred. No. 4.2e-10;
Matches 212; Conservative 0; Mismatches 214; Indels 9; Gaps 1;

QY 124 ATATCCCTTAAATGCAAGTTGACATGACATGCGCTGCTGAGTATGCGCTGC 183
DB 20 ATTTTCAATTAAGTAAGAAAGTAGCTTAATTCAGAGTCTGTTTATGGGATTTGGCTTT 79
QY 184 AATGTCAAAAACGCTGCTGATGACAGAGCAAAAGTACTATTGATTTGACCGGAAGCG 243

Db 80 AGATTGCCAANTCCCTAGCTGAAGCAGACACAAATTGTTTTTAATCACTTATCTCTG 139
QY 244 ACAAACTGCACAAGATTGTGCTGAGTT-----AGCGCAAAACCGCTACGCGCTG 294
Db 140 AATCAGTTGAACAAGCTTGGAAATTAATGCGCGCAGGAGATTGAACAGACTGCTATG 199
QY 295 AACTGATCTCTTCAATATCAGCACTGCATTAACATGCTGGCGGACATTAATGCACTG 354
Db 200 TGTGTGATGTCTACATGAAGAGCAAGTCAAGCGATGTGCTCAATTAAGAAAG 259
QY 355 CGGCTGAGCTGATATTTTTCATGCGCAATGAGCGCTTATTTGGGCGCCAGTGTGCTG 414
Db 260 TCGGCTTATTTGATATTTTGTCAATATATGCGGAAATTAATGAATTCATGCTG 319
QY 415 AAGGTATCCAGATGTCGTGGAGTCTGTTAAATCTGAATTAATGCGCGCTTTCGCT 474
Db 320 ATATGTGACGAGAAGATTCCTGCAAGTGAATGATGTTGACTGTAATGCGCATTTTATTA 379
QY 475 GTGTCCGTGAGATCTCTCCCATATGATTCGCGCAGAGGTGGCGCATTAATTTTACCA 534
Db 380 TGGCAAAAGCGGTATTCAGATATGATGAAGGCTCATGCGCAAAATTAATCAATATCT 439
QY 535 GTTCATCGCGGCGG 549
Db 440 GTTCATGATGAGTG 454

RESULT 6
US-09-070-927A-238
Sequence 238, Application US/09070927A
Patent No. US20020120116A1

GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 238:
SEQUENCE CHARACTERISTICS:

LENGTH: 29729 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 238:
US-09-070-927A-238

Query Match 1.6%; Score 64.6; DB 10; Length 29729;
Best Local Similarity 48.7%; Pred. No. 7.3e-09;
Matches 212; Conservative 0; Mismatches 214; Indels 9; Gaps 1;

QY 124 ATACTCCCTTAATGCGAAAGTTGACACTATCACTGCGGCTGCTGAGTATGGCTGC 183
Db 15623 ATTTTCAATTAGTGAAGAGTGAAGCTTAATTAACAGTGTCTGTTATGGATTGGCTTGG 15682
QY 184 AATGTGCAAAAGCGCTCTGATGACAGACCAAGTAGTATGATTGACCAGGAGCG 243
Db 15683 AGATTGGAAATCCCTGAGCAAGCAGACCAACAAATTGTTTTTAATCACTTATCTCTG 15742
QY 244 ACAAACTGCACAAGATTGTGCTGAGTT-----AGCGCAAAACCGCTACGCGCTG 294
Db 15743 AATCAGTTGAACAAGCTTGGAAATTAATGCGCGCAGGAGATTGAACAGACTGCTATG 15802
QY 295 AACTGATCTCTTCAATATCAGCACTGCATTAACATGCTGGCGGACATTAATGCACTG 354
Db 15803 TGTGTGATGTCTACATGAAGAGCAAGTCAAGCGATGTGCTCAATTAAGAAAGAG 15862
QY 355 CGGCTGAGCTGATATTTTTCATGCGCAATGAGCGCTTATTTGGCGGCCAGTGTGCTG 414
Db 15863 TCGGCTTATTTGATATTTTGTCAATATATGCGGAAATTAATGAACGATTCATGCTG 15922
QY 415 AAGGTATCCAGATGTCGTGGAGTCTGTTAAATCTGAATTAATGCGCGCTTTCGCT 474
Db 15923 ATATGTGACGAGAAGATTCCTGCAAGTGAATGATGTTGACTGTAATGCGCATTTTATTA 15982
QY 475 GTGTCCGTGAGATCTCTCCCATATGATTCGCGCAGAGGTGGCGCATTAATTTTACCA 534
Db 15983 TGGCAAAAGCGGTATTCAGATATGATGAAGGCTCATGCGCAAAATTAATCAATATCT 16042
QY 535 GTTCATCGCGGCGG 549
Db 16043 GTTCATGATGAGTG 16057

RESULT 7
US-09-974-300-2287
Sequence 2287, Application US/09974300
Patent No. US20020146721A1

GENERAL INFORMATION:
APPLICANT: Berka, Randy M.

APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods For Monitoring Multiple Gene

FILE REFERENCE: Expression

CURRENT APPLICATION NUMBER: US/09/974,300

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: 09/680,598

PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279,526

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 8481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2287

LENGTH: 468

TYPE: DNA

ORGANISM: Bacillus licheniformis

US-09-974-300-2287

Query Match 1.6%; Score 63.4; DB 10; Length 468;
Best Local Similarity 46.8%; Pred. No. 6.9e-10;
Matches 199; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 174 ATGGCCTGCAATGTGCAAAACGCTGCTGATGACAGACCAAGTAGTATGATTGAC 233
Db 34 ATGGCCTGCAATGTGCAAAACGCTGCTGATGACAGACCAAGTAGTATGATTGAC 233
QY 234 CGGAAAGCGAACAACCTGCAAGATTTGCTGAGTTAGGCGAAAGCGCTACGCGCTG 293

Db 94 GCGAAGCAGCAGAGTCGAGACGAATCAAGCGCTGGCGCGCATGGCTTTT 153
Oy 294 CAACGTGATCTCTCAATATATGCAAGTCGATTAACATGCTGGCGACATTAATGCACTG 353
Db 154 AAAGCGAGCTTCCATCGCGATGAGTTCAAGCGCATGATGAAGAGCGGCGACCG 213
Oy 354 GCGGCGGCTGGATATTTTTCATGCAATGACGCGCTTATATTTGGCGGCCAGTGGCT 413
Db 214 TTGCGACCGCTTACATCTCTTGTCAACCAATGGCGCATTAACATCAATCTGTTCAAT 273
Oy 414 GAAGTGATTCAGATCTCGGATCGCTGTTAAATCTGAATTAATGCGCGCTTTCGC 473
Db 274 AGAATGAAGAGATGAATGGGACGACGTCATTAACATTAAGTTAAAGTGTTCAT 333
Oy 474 TGTGTCGTCGATCTCTGCGCATATGATTTGGCGACAGTGGCGCATATATTTTACC 533
Db 334 TGTTCAAAAGCTGTGACAGACAGATGATGAACAAAGACGCGCGCATCAATATC 393
Oy 534 AGTTCATCGCGCGCTCGTTCGGTTATCTGGGAACCGATCAACCGCGTCAATTT 593
Db 394 ACCTCGGTGTAGCGCTCTGTCGTAACGCCGCGACGCACTATGTGCGGCTAAATCA 453
Oy 594 GCGGT 598
Db 454 GCGGT 458

RESULT 8

US-09-974-300-6558
; Sequence 6558, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6558
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6558

Query Match 1.6%; Score 63.2; DB 10; Length 744;
Best Local Similarity 45.4%; Pred. No. 1.1e-09;
Matches 227; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Oy 267 GAGTTAGCGCAAAACGCTGACGTCGACATCTCTCATATATATGCAAGTCGAT 326
Db 151 GCGCTTGGCCCAAGAGCGCTTGTGCTATCAGGCTGACGTTCTCTCAAAAGGAGTCGCA 210
Oy 327 AACATCTGCGCGACATTTATGCACTGCGCGGTGGCGTGGATATTTTTCATGCAATGCA 386
Db 211 GCGATGATGAAGAGAGATGGCGCTTCAATCAATGATATATTTTATGAGCAATGCC 270
Oy 387 GGGGCTTATTTGGCGGCCAGTGGCTGAAGTGAATGCAATGTCGGATCTGTGTTA 446
Db 271 GGGATTAACGGCGATTAATGCTGATGAGATGAAGAAAGACGATTGGGATGCGCATTT 330
Oy 447 AATCTGAATTAATGCGCGCTTGTGCTGTCGTCGACATCTCTGCGCGCATATGATTCGCG 506
Db 331 GATACCACTTAAAGAGGCTTCCATTTGCGCGCAAAAGCACTACGCCGCAATATGATGAAA 390
Oy 507 CAGAGTGGCGGATATATTTTACAGTTCCATGCGCGCGCTGCTTCCGTTATCTGG 566

Db 391 CAACGTGTCGCGCAAGATATCATATGTCGTCTGTGCTTGGCGCTAATGGGTACGCTGGG 450
Oy 567 GAACCATCTACACCGCGCTCAATATTTGCCGTCAGGCAATGCTATACACTACCGCGCG 626
Db 451 CAAGCGAATTAATGTCCTCCGCAAGAGAGTGTATGCTGATGACTTAAGTCTTTGGCGG 510
Oy 627 CAGTTTCTCAATATATGCGCGTGTGGGTGGGCTGCTGTCGCGCAGACAGTACACTGCC 686
Db 511 GAATTTGCGAGGCGAGCATTTCTTTCATGCGGTGCGCGCGCTTTATTAACAGAT 570
Oy 687 CTGCTGATGACTGGCCAAAGCCAAATGAGAAAGCCCTGCGCAATGATGCTGATG 746
Db 571 ATGACAGATGAGCTACCAAGCGAAGCAAGACAGACGTGTTAAACAAATCCATTGGGG 630
Oy 747 CAACCATTAAGTGGCGGA 766
Db 631 AAGCTTGTGAACCGAGGA 650

RESULT 9

US-09-712-363-99
; Sequence 99, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-99

Query Match 1.5%; Score 58; DB 9; Length 807;
Best Local Similarity 48.8%; Pred. No. 6.3e-08;
Matches 157; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Oy 378 GCCAATGCGCGCTTATATGCGCGCCAGTGGATGAGTGCATGCTGATG 437
Db 271 GCCAAGCGGGTACCGCGCATTCGCGCGATGCTGCTGCTTGGCGGAAAG 330
Oy 438 CGTGTGTTAATGATATATATGCGCGCTTGTGCTGCTGCGTGCAGTCTGCGCAT 497
Db 331 ACCGAGGTGAGTTGATATGCGCGTGCAGACCTTACGTTAGTGGGTGTCGCGGCG 390
Oy 498 ATGATGCGCAGAGTGGCGATATATTTTACAGTTTCATGCGCGCGCTGTTCCG 557

Db 391 ATGATGAGCGCAAGCGCGCGCATCTTGATTTCTGTCGCGCGCGCAATTCACCG 450
Qy 558 GTTATCTGGAGACGATCTACACCGCGCTCCAAATTTCCGTTGAGCATTCGTACACACT 617
Db 451 ATTCCCTACAAACGCGACCTATGCGCGACCAAGCCCTTCGTGAACACCTTCAGGAATCT 510
Qy 618 ACCGCGCGCAAGTTTCTCAATATGCGCGCTGCGTGGTGGTCCAGACAGTA 677
Db 511 CTGGCGCGTACGATACCGCGCTCGCGCTGACGATGCTGCTCGCGCGCGCGCT 570
Qy 678 GTACCTGCGCTGATGACT 699
Db 571 CGCAGCGATACCGGATGCT 592

RESULT 10
US-09-813-453A-71
Sequence 71, Application US/09813453A
Patent No. US2002016861A1
GENERAL INFORMATION:-
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: CG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 71
LENGTH: 8320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: plasmid, pAN296
US-09-813-453A-71

Query Match 1.4%; Score 56.4; DB 9; Length 8320;
Best Local Similarity 47.7%; Pred. No. 1.3e-06;
Matches 165; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
Qy 347 CGAAGTGGCGGCGCTGCGATATTTTCATGCCATGACGCGCTTATTTGGCGGCC 406
Db 639 GATTCAGATCGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 698
Qy 407 AGTGGCTGAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 466
Db 699 GGTGTTGACTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 758
Qy 467 GTTTCGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 526
Db 759 GATTCGCTGTAACAAACGCGGCTTCGCAAAAGCTTGACCAAAAGGACATTCAT 818
Qy 527 TTTTACAGTTCCATCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 586
Db 819 CAATATGCGCTCTCAAGCGGGAATGCGCACACCAATCTAGCTGTATTCGCGAC 878
Qy 587 CAATTTGCGCTTACAGCATTCGACACTACCGCGCGCGAGTTTCTCAATATGCGCT 646
Db 879 CAATATGCGCTTACAGCATTCGACACTACCGCGCGCGAGTTTCTCAATATGCGCT 938
Qy 647 GCGTGGGCGGCTGCTGCGACAGCATGATGATGATGATGATGATGATGATGATG 692
Db 939 TTAATGCAACAGTCAACCGCGCGCGATTCAGACGACTTTT 984

RESULT 11
US-09-070-927A-4/c
Sequence 4, Application US/09070927A

Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6948 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-070-927A-4

Query Match 1.4%; Score 55.6; DB 10; Length 6948;
Best Local Similarity 46.4%; Pred. No. 2.1e-06;
Matches 218; Conservative 0; Mismatches 249; Indels 3; Gaps 1;
Qy 141 AAGTTGCACTATCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 200
Db 3137 AAGTTATTTGTTATCATGCGGCTTCAGATGCGATGCGTGAAGCAACAGCCCTTACT 3078
Qy 201 CTGATGACGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 260
Db 3077 GCCAGAAAGAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGAT 3018
Qy 261 GTGCGTGAAGTAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 317
Db 3017 AAAAAAGAACTCTCTGAGCAAGATTCGTGTCACAAACAGATGATGATGATGATG 2958
Qy 318 CAAGTCGATTAACATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 377
Db 2957 GAAGTTACGCGTATTAATCAAGCTTACATGGAATAATATGAGAGATGATGATGAT 2898
Qy 378 GCGATGACGCGCTTATATTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 437
Db 2897 AACAAATGACAGAGTATGCAACGCGCGCTTATGAAACACCAAAAGAGAAATGCG 2838
Qy 438 CGTGTGTTAATCTGAATATTAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 497
Db 2837 CAATGCTAGATATTAATATGAGGTTTAAATGCGATGCGGCGGCGGCGGCGGCGG 2778


```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (250)..(1185)
US-09-885-303A-13

```

Query Match	1.38;	Score 50.6;	DB 9;	Length 1714;
Best Local Similarity	47.1%;	Pred. NO. 3.1e-05;		
Matches 155;	Conservative 0;	Mismatches 174;	Indels 0;	Gaps 0;

Oy	336	GGGTCGGCCGATATTTTTCATATGCCAATGACAGGCGCTTAATATTTGGCCGACGAGCGCTGA	415
Db	492	GGGAGAACTGGACGTCGTCGTGATTAATATGCTGGAAATGGCGCTGGTGGGGCCCTCGAGAGG	551
Oy	416	AGGTGATCCAATATGTCGTGGATTCGTGTAAATCTGTAATTAATATGGCGCTTTCGCTG	475
Db	552	GCTCAGCCTTCTGTCGCATGACGAATATGCTTTTGACACCAACTTTTTCGGAGCTGTCCGCTCT	611
Oy	476	TGTCGTCGTCAGTCCGTGCGCATATGATTTGCGCAGAGTCGCGGCATTAATATTTTACAG	535
Db	612	CCTCAAACTGTGCTGTCCAGGCATGAAGAGGAGCGGAGGCCACATCTGTCGTGATACG	671
Oy	536	TTCATATCCCGGGCGTCGTCGCTTATCTGGGAACCGATCTACACCGCGTCCAAATTTGC	595
Db	672	CAGTCTCAATGGAGCCCTGCAGGCTGTCAATCTTCAACGATGTCATACCACTTCCAAAGTTTCGC	731
Oy	596	CGTTACAGCATTTCTGTACACACTACCCCGCCGAGCTTTCATTAATATGGCGTCGTGGG	655
Db	732	CCTGGAGGATTTCTCGAAACCTTCGTCATATCCACACTGTGTGCAAGTTCAACATCTTCATCTC	791
Oy	656	TGCGGTCTGCAGAGACCAGTATGACATCG	684
Db	792	CCTGTTGAGCGCAGCCCGCTGGTCCACGG	820

```

US-RESULT 15
US-09-974-300-2280
: Sequence 2280, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy M.
: TITLE OF INVENTION: Methods For Monitoring Multiple Gene
: FILE REFERENCE: 10085, 500-US
: CURRENT APPLICATION NUMBER: US/09/974, 300
: CURRENT FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: 09/680, 598
: PRIOR FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/279, 526
: PRIOR FILING DATE: 2001-03-27
: NUMBER OF SEQ ID NOS: 8481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2280
: LENGTH: 680
: TYPE: DNA
: ORGANISM: Bacillus licheniformis
US-09-974-300-2280

```

Query Match	1.3%	Score 50.2;	DB 10;	Length 680;
Best Local Similarity	52.1%	Pred. No. 2e-05;		
Matches 112;	Conservative	0;	Mismatches 103;	Indels 0;

Oy	334	TGGCGGCAATATATGCAATGGCGGGGGCGGATATTTTTCATGCCAATGACGGCTT	392
Oy	218	TGCGTAGGCAAAACACGCGGTTTGGCGGATCGATATTTTGGTCAACAATGCGCGCTTA	277
Oy	394	ATATTGGCGGCGCCAGTGGCTCAAGAGTGATGTCAGATGTCGTGGGATCGTGTTAAATCGA	453
Oy	278	ATTTGGCGCAAAACGACCGCTTATACGTGCTTGAAGAAATGAAAGATGATTAATTTGA	333
Oy	454	ATATTAATGCGGCGCTTTGCTGTGTGCGCTGACAGTCTGCCCATATATATGGCAGAGGT	513
Oy	338	ATTTAACCGGACCTTTTCTATGTGTACAGGCTGTCAATCCGGAAATGATACAGGAGGCG	397

Oy 514 CGGGCGAATAATTTTACCACTCCATCGGGGC 548
||| ||| ||| ||| |||
Db 398 GCGGAAATTCGTGAACATGACTCTAGCAGCAGC 432

```
Search completed: March 21, 2003, 15:43:42
Job time : 380.853 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:18:35 ; Search time 6094.69 Seconds
(without alignments)
10552.182 Million cell updates/sec

Title: US-09-802-208B-2

Perfect score: 3971

Sequence: 1 atcagctgagcagcttgcgtc.....tcacgcgcgtcagcagcat 3971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199.4	5.0	692	13	BI260394
2	189.2	4.8	922	14	BQ934625
3	185.8	4.7	926	14	BQ644899
4	171.4	4.3	736	13	BI759992
5	169.8	4.3	905	14	BQ440371
6	169.8	4.3	1069	14	BM921107

7	156	3.9	676	14	BM791109	BM791109 K-EST0071
8	153.8	3.9	535	17	AZ302589	AZ302589 GSSBR152
9	150	3.8	655	17	AZ650403	AZ650403 1M05201.01
10	150	3.8	712	13	BI557944	BI557944 603237561
11	149.2	3.7	714	13	BI149014	BI149014 602911220
12	148.4	3.7	503	12	BF440790	BF440790 256813 MA
13	147.8	3.7	676	9	AI527876	AI527876 u130C03.Y
14	144.4	3.6	670	13	BM646051	BM646051 170006873
15	142.4	3.6	983	14	BQ932942	BQ932942 AGENCOURT
16	140.6	3.5	699	12	BF783891	BF783891 602109920
17	139	3.5	853	12	BF218299	BF218299 60181643
18	137.6	3.5	796	12	BF6768597	BF6768597 602741908
19	135.8	3.4	654	9	AI108240	AI108240 GH07026.5
20	135.4	3.4	666	9	AI388552	AI388552 GH19496.5
21	135	3.4	857	13	BI329415	BI329415 602980145
22	134	3.4	479	9	AI196200	AI196200 u170D11.Y
23	132.8	3.3	669	14	BQ406350	BQ406350 GA_Ed009
24	131.4	3.3	658	13	BI629061	BI629061 RH57864.5
25	130.8	3.3	621	10	AM164681	AM164681 se76b09.Y
26	123.2	3.1	602	9	AI789136	AI789136 uK51d10.Y
27	122.2	3.1	1079	13	BI819693	BI819693 603037856
28	120	3.0	533	9	AL749866	AL749866 AL749866
29	117.8	3.0	1053	11	AK09249	AK09249 Mus muscu
30	117.4	3.0	730	9	AL529279	AL529279 AL529279
31	117.4	3.0	851	14	BQ892439	BQ892439 AGENCOURT
32	116.2	2.9	637	13	BI565802	BI565802 RH63626.5
33	115.6	2.9	634	10	AM107094	AM107094 u192h06.Y
34	114.4	2.9	536	13	BM036183	BM036183 ful6h06.Y
35	112.8	2.8	632	13	BI613997	BI613997 RH43270.5
36	111.8	2.8	715	14	BQ801483	BQ801483 WHE2814_H
37	111.2	2.8	604	13	BI588685	BI588685 RH3085.5
38	111.2	2.8	612	13	BI57185	BI57185 RET0835.5
39	109.6	2.8	386	14	BM855062	BM855062 K-EST0137
40	109.6	2.8	907	12	BC184645	BC184645 RST7307 A
41	107.6	2.7	1095	17	CNS06C90	AL392248 T3 end of
42	105.6	2.7	531	10	BE682075	BE682075 179972 MA
43	104.4	2.6	582	13	BI583725	BI583725 RH22404.5
44	104.2	2.6	486	14	BM856746	BM856746 K-EST0140
45	104.2	2.6	795	9	AU121672	AU121672 AU121672

ALIGNMENTS

RESULT 1
BI260394
LOCUS 602969456F1 NIH_MGC_12 Homo sapiens CDNA clone IMAGE:5108988 5',
DEFINITION mRNA sequence.
ACCESSION BI260394
VERSION BI260394.1 GI:14818651
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 692)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue/Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM11263 row: p column: 13
High quality sequence stop: 692.
Location/Qualifiers
1. 692

FEATURES
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5108988"
/clone_lib="NIH_MGC.12"
/tissue_type="cervical carcinoma cell line"
/ab_host="DH10B"
/notes="Organ: cervix; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT      171 a      149 c      204 g      168 t
ORIGIN

Query Match      5.0%; Score 199.4; DB 13; Length 692;
Best Local Similarity 60.9%; Pred. No. 9.6e-48;
Matches 361; Conservative 0; Mismatches 226; Indels 6; Gaps 2;

QY 1086 AGCATCGGTTTGATGCCACCTGTTCTGTGGTACTGATTAATAACGCTGATCCATT 1145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25 AGACACTTGGGTTTGATGCCACGCTGTTCTGTGGTATGAGCATGTTCCACCATTT 84

QY 1146 GCCTGTCAGCCCGGAGAGATGCAAAATATCATTTGTGTGATGATCACCGCGC 1205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85 ACCAGTCAACCGAAGAGGATGCCATGCAACGTCATCATGTGGCTGGACCATGAGC 144

QY 1206 CACCGAACAAGCAGAGGAATCAATGCCACTCACCATCCGCTGTAAGTACGTGGTGG 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 145 AGTCAGTCAAGTTTAAACAGGATCAATGAGACCAAGCAGTCTCCACGTAAGCTGGGGG 204

QY 1266 TAAATTTTCGCTGTAATGGAACACCGAATAATTCCTGGCTGAAAGAAATATGCGAGA 1325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 GGTGATGCTCTGTGAAATGACAGAGCCCGAATCTTGTGCTGAAAGAAAGATTTGAGAGA 264

QY 1326 GAT--CTACGAACGTCGCCGACAAATTTTTCGATCTGGCCGATTTTTCGACTGGCGGC 1382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 265 GATTTCGTGGATTAAGCGGAGCATTTCTTGATCTCCAGACTTCTTATTCGTGGAGAGGC 324

QY 1383 TACCGGTGATTTAGCGGCTCAGATGACATGTTACTGTAAATGACAGTGGCTGGCACA 1442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 325 AACAGGCTGACAGCAGCAGGCTCTGCTGCTCCGTTGTAAGTGAATATTCAGCA-- 382

QY 1443 TGAATATCGCTGGGATCCAGATTTATTTCCGACCATTTGGCCCTTCGACAGTTAGCGGATGA 1502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 383 -GAGAAAGCGCTGGGACAGCAGTTTCTGAAATGATTTGTTGAAACACTTTGTTGACAGA 441

QY 1503 AGATTTTATTCGATGCTCATATATTTGTTCTCCGGAACACCTGTTGGAATGAGTTT 1562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 442 TAATTACAGCAAAATAGAAACCAAGTGTACTCTCTGAGACTTCTCTTGAATATGGGCT 501

QY 1563 AACAGCACAAGCCGCGGACAGATGGATTTACTCCCGGACACACTTCGCTGTAGAGGTT 1622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 502 CACACCGAGGACAGCAAGAACCTTGGCTTCTCCCTGGGATTTGCGGCTCCAGCTTCACT 561

QY 1623 AATTGATGCTCAGCGTGTGGCATCGGTGCGGTGAAGAGTGGAGCGC 1675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 562 CATTGATGCCCAATGACAGAGGACTAGAGATGATTTGGGCGCAGATGTGAGAGGGC 614

RESULT 2
BOJ34625      922 bp      mRNA      linear      EST 21-AUG-2002
LOCUS      BOJ34625
DEFINITION      AGENCOURT_8818058 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6422149
5', mRNA sequence.
ACCESSION      BOJ34625
VERSION      BOJ34625.1 GI:22350008
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 922)
AUTHORS      NIH-MGC http://mgc.ncl.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
JOURNAL
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DFP/Gazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2598 row: k column: 14
High quality sequence stop: 723.

FEATURES
source
location/Qualifiers
1..922
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6422149"
/clone_lib="NIH_MGC.18"
/tissue_type="large cell carcinoma"
/ab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC library."

BASE COUNT      220 a      205 c      275 g      218 t      4 others
ORIGIN

Query Match      4.8%; Score 189.2; DB 14; Length 922;
Best Local Similarity 56.4%; Pred. No. 1.3e-44;
Matches 433; Conservative 0; Mismatches 322; Indels 13; Gaps 4;

QY 879 TATGTTGTTGATGTTGGATCAGCAGTGTCCGCGCGGATTTTGTATGTCACAGGATC 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 TGTGGTGTGACCTTGGAAACAGCAGTGTCCGTCAGCTCTGTGTGACCAAGATGGGGT 222

QY 939 TCTGCTATCCCATGCCACAGAAAAATCAGACTACGCGCGCAGCGAAAGCCGCTGGA 998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 CTTGTGGCTTTTTCACACACAGCCCAATTAAGATTTGGAGGCCCAAGTTCAACCAACATGA 282

QY 999 ACACTCCAGCAGAGATCTGGCAGCGGCTGTCTTCAATTAATGTAATGCGCTCACTCT 1058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 283 GCACTCTCCGAGACATCTGGGCTGCTGTGTGTACAAAGAAAGT-----TGT 336

QY 1059 GGCAGAGCTTTGTGCACAAAGTGTGGAGGAGCATGTTTGTGATGCCACTGTTCTGTGT 1118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 337 ACAAGGATTTGATTTAAACCAATTCAGAGACTTGGTTGTATGCTCCAGTGTCTGTGT 396

QY 1119 GGTACTGATTAATAACCGTGTATTCCTGTACAGCCCGAAGAGATGCANAAGCAAA 1178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 397 TGTTTGGATTAAGAGATTTCAACCATTAACAGTCAACAGGAAGGAGATTCCTTCGAAA 456

QY 1179 TATCATTTGTGTGATGATCAAGCGCGCACCAAGCAGAGGATTCATGCTCACTCA 1238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 457 CGTATCATGTGGCTGGACCATCGACAGTCACTGAAGTTAAACAGATTCATATGAGACAA 516

QY 1239 CCATCCGCTGTGAACAGTGGTGTAAATTTCCCTGAAATGGAACACCGAATAAT 1298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 517 GCACAGTGTCTCTCAGTACTCGGGGGGGTGTATCTCTGTGAAATTCAGAGGCCCGAATACT 576

QY 1299 TCTCTGCTGAAGAAATATGCGCAGATCTAC--GAACGTGCGGACGAATTTTTCGA 1355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 577 TCTGTGGCTGAAGAGAACTTGAAGAGATTTGTGGATTAAGCGGGACATTTCTTTGA 636

QY 1356 TCTGGCCGATTTTCTGACCTGGCGGCTACCGGTGATTTAGCGGCTTCAGTATGACTGT 1415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 637 TCTCCCGACTTTCTTATTCGGAAGGATGTGCACAGCGAGCTCTCTCTCTCTCTCT 696

QY 1416 TACTCTTAATGACGCTGGCTGGACATGAATAATCGTGGGATTCAGATATTTTCCGAC 1475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db	697	GGTGGTAAAGTGGACATATTCAGCA--NAGAAAGCGCTGGGACGACAACTTTCTGGAAAT	753
QY	1476	CATTGGCCCTTCAGAGATTAGCGGATGAAGATTTTATTCGCAATGGTCAATCATATTTGTTTC	1535
Db	754	GATTGGTTTGGAAAGCACTTTGTTGCAGATAATTTACAGCANNATAGG-AAACAAGTCTACCC	812
QY	1536	TCCCGGAACACCTGTTGGAAATGGTTTAACGACCAACACCGCGCAGATGGGCAATTACT	1595
Db	813	TTCTGGACACTTCTCTTGGAAATAGGGCTCACACGACGACGACGAAGAAGACCTTGGCCCTTCTC	872
QY	1596	CCCCGGCACACCTGTGCGCTGAGGGTGAATGATGCTCAGCTGGTGG	1643
Db	873	CCGTGGGATTGGCGGTGCGACCTTCACATATTGATGCCATGGCGGGG	920
RESULT 3			
LOCUS	BO644899		
DEFINITION	BO644899	926 bp	mRNA linear EST 15-JUL-2002
ACCESSION	AGENCOURT_8493114	NIH_MGC_100	Homo sapiens cDNA clone IMAGE:6300239
VERSION	BO644899		
KEYWORDS	BO644899.1	GI:21769071	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	1 (bases 1 to 926)		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ .		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		

found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MCC clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1CM251d row: c column: 24
 High quality sequence set: 546.

FEATURES	Location/Qualifiers
source	1. .926

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:630029"
/clone_id="NH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notice="Organ. livery. Vector: pORB7. Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGCGACGG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NH_MGC
library."

```

BASE COUNT	222 a	213 c	267 g	223 t	1 others
ORIGIN					
Query Match	4.7%	Score 185.8;	DB 14;	Length 926;	
Best Local Similarity	60.2%;	Pred. NO. 1.4e-43;			
Matches 345;	Conservative	0;	Mismatches 222;	Indels 6;	Gaps 2

QY 1106 CTTTCTTCCTGAGGACATGATAAAAAACGGGATCATTCATGGCTCTGACGCCGGAGAG 1165
 Db 1 CGTCTTCCTGAGGATTTTGGATTAACACAGTTTCACCATTTACCACTGACCAACGAGAAAGCG 60
 QY 1166 ATGCMAAGCAAAATATCATTTGTGTGATGATGATCCGGCCACCGACAGCAGACGAGCA 1225
 Db 61 ATTCCCATGCAAAACGTCATCATCTGTGGCTGTGACCATGTGACACAGTCACAGTAACTTAACAGGA 120

OY	1226	TCATGCCACATCACCATCCGGGCGTGAACATACGTGGGAGTAAATTTGCGCTGAAATGG	1285
Db	121	TCATAGAGACCAAGCACAGTGTCTCCACAGTACGTGGGGGGGTGATGTCGTGGAAATGC	180
OY	1286	AAACACCGAAAAATTCCTGCGGTGAAGAAAGAAATATGCCAGAGAT---CTACGAACTGGCG	1342
Db	181	AGGGCCCGAAACTTGTGTGGCTGAAAGAAACTTGAGAGAGATTTGTGGGTAAAGCGGG	240
OY	1343	GACAAATTTTTCGATCTGGCCGCAATTTTTCGACCTGGCGGGCTACCGGTGATTAAGCGGTT	1402
Db	241	GACATTTTTCGATCTCCCGGACATTCATCTGTGAAGGCAACAGGTGCACAGGACGAT	300
OY	1403	CAGTATGCACATGTACTGTAAATGAGAGTGGCTGCGACATGAAATATCGCTGGATTCAG	1462
Db	301	CTCTCTGCTCCCTGGTGTGTATAGTGGACATATTACACA---GAGAAAGGCTGGGACGACA	357
OY	1463	ATTATTTCCGACACATTTGGCCTTGAGAGATTAGCGATGAAAGATTTTATTCGATTTGGTC	1522
Db	358	GTTTCTGGAAATATGATTTGGTTTGAAGACTTTGTGCAGATTAATTACGCAAAATATGANA	417
OY	1523	ATCATATTTGTTCTCCCGGAACACCTTGTGGAAATGCTTTAACACACAGCGCGGGCGAG	1582
Db	418	ACCAAGTCTCACTCTCTGAGCTTCTCTTGGAAATGGGCTCAACACAGGAGGAGAAAGG	477
OY	1583	AGATGGGATTTACTCCCGCGACACCTCTGCGTGTAGGGTTAATGATGCTACGCTGGTG	1642
Db	478	ACCTTGGGCTTCCCTTGGGATTTGGGTGCGCAGCTTCACTATTATGCCATTCGACGAG	537
OY	1643	GCATCGTACGGTTGGCTAGACGTTGACGCC	1675
Db	538	GACTAGGAGTATTTGGGACAGATTTGAGAGGGC	570

RESULT 4	
B175992	
LOCUS	B175992 736 bp mRNA linear EST-25-SEP-2001
DEFINITION	603044443f1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5184755', mRNA sequence.

ACCESSION	BI759992	
VERSION	BI759992.1	GI:15751570
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (phases 1 to 736)	NIH-MGC	http://mgc.ncl.nih.gov/ .		
National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
Contact: Robert Strausberg, Ph.D.				

```
Email: cgapbs.r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1461 row= e column= 12
High quality sequence stop= 736.
      | Location/Qualifiers
```

```

source
1. .736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5184755"
/clone_1ib="NTH.MGC.116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach. Vector:
pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is

```



```

ACCESSION   BM921107
VERSION     BM921107.1  GI:19371486
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 1069)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            JOURNAL   Unpublished (1999)
            TITLE     Unpublished (1999)
            COMMENT   Email: c9apbs-remail.nih.gov
                    Tissue Procurement: Life Technologies, Inc.
                    cDNA Library Preparation: Life Technologies, Inc.
                    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                    DNA Sequencing by: Agencourt Bioscience Corporation
                    Clone distribution: MGC clone distribution information can be
                    found through the I.M.A.G.E. Consortium/LLNL at:
                    http://image.llnl.gov
                    Plate: L1AM12786 row: k column: 06
                    High quality sequence stop: 726.
                    Location/Qualifiers
                        1..1069
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:5752445"
                        /clone_1lb="NIH_MGC_115"
                        /lab_host="DH10B"
                        /note="Organ: pooled brain, lung, testis; Vector:
                        PCMV-SpORF6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                        source anonymous pool of 6 male brains, age range 23-27; 1
                        male lung, age 27; and 1 male testis, age 69. Library is
                        oligo-dT primed and directionally cloned (EcoRV site is
                        destroyed upon cloning). Average insert size 1.8 Kb,
                        insert size range 1-3 Kb. Library is normalized and
                        enriched for full-length clones and was constructed by C.
                        Gruber (Invitrogen). Research Genetics tracking code
                        021. Note: this is a NIH-MGC Library."

BASE COUNT   233 a 270 c 304 g 262 t
ORIGIN
Query Match      4.3%; Score 169.8; DB 14; Length 1069;
Best Local Similarity 58.7%; Pred. No. 9.8e-39;
Matches 335; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

QY 879 TATTGCTGTGATGTTGGGATCAGCAGTCTCCGCCGGGATTTTGAATCAACGATC 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 227 TGTGGGTGTGGACGTTGGAAACGCTGCTCCGTGACCTGTGTGACAGAGTGGGT 286
QY 939 TCTGATCCATGCCACAGAAATATCAGACTACGCGCGCAGCGGAAGCCGCTGA 998
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 287 CCTGTGGCTTTTGCAGACCAACCAATTGAAGATTGGAGCCCCCACTTAACCACTATA 346
QY 999 ACAGTCCAGCAGAGACTGTGCGAGCGGCTGTTCTATGTTATGCGCTCACTCT 1058
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 347 GCAGTCTCCGAGACATCTGGGCTCGCTGTGTGTGCACAAGCAAGT-----TGT 400
QY 1059 GGCAGACGTTTGCACAAGTGTGCGAGGATCGTTTGAATGCCACCTGTTCTGTGT 1118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 401 ACAAGGATGATTTAAACCAATTCGAGACTTGGGTTGATGCCACGTTCTTGTGT 460
QY 1119 GGTACTGATTAATAACGATGATTCATGCTGTGACCGCGGAAGAGATCAAGCAAA 1178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 461 TGTTTTGAATGAAGATTACACCAATTAACCAAGCAAGAGGATTTCCCATGAAA 520
QY 1179 TATCATTTGTGTGATGATCAACCGCGCACGAACAAGAGGATCAATGCCACTGA 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 521 CGTCAATCATGTGGCTGACCATGACAGCTCAAGTTAAACAGAGATCAATGAGACAA 580
QY 1239 CCATCCGGTGTGACTAGTGTGGTGTAAATTTTGGCTGAATAGGAACACCGAAAT 1298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 581 GCACAGTGTCTTCAGATACGTGGGGGGGTGATGTGTGGAATTCAGAGCCCGCAACT 640

```

```

QY 1299 TCTCTGCTGAAGAAATATGCGACAGAT---CTACGACGTCGCCGACAAATTTTGA 1355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 641 TCTGTGGCTGAAGAGACATTGACAGAGATTGCTGGGATTAAGCGCGGACATTTCTTTGA 700
QY 1356 TCTGGCGGATTTTCTGACCTGGCGGCGCTACCGGTGATTTAGCGGCTTCAGTATGCACTGT 1415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 701 TCTCCCGGACCTTCTTATTCGTGGAAGCAACAGTGTGCACAGCAGGCTCTCTGCTCCCT 760
QY 1416 TACCTGTAAATGACGTGCTGGCAGATGAA 1446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 761 GGTGTGTAAAGTACATATTCAGACAGAGAAA 791

RESULT 7
BM791109
LOCUS
DEFINITION K-EST0071070 S21SNUS20 Homo sapiens cDNA clone S21SNUS20-19-B10 5',
            mRNA sequence.
ACCESSION   BM791109.1  GI:19139341
VERSION     BM791109
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
            1 (bases 1 to 676)
REFERENCE   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
            21C Frontier Korean EST Project 2001
            Unpublished (2002)
            CONTACT: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@email.kribb.re.kr
            Plate: 19; row: B column: 10
            High quality sequence stop: 676.
            Location/Qualifiers
                1..676
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="S21SNUS20-19-B10"
                /clone_1lb="S21SNUS20"
                /sex="F"
                /tissue_type="Stomach"
                /cell_line="SNU-520"
                /lab_host="Top10F"

TITLE
JOURNAL
COMMENT

```

```

BASE COUNT   166 a 143 c 204 g 163 t
ORIGIN
Query Match      3.9%; Score 156; DB 14; Length 676;
Best Local Similarity 58.8%; Pred. No. 9.8e-35;

```

Matches	328	Conservative	0	Mismatches	220	Indels	10	Gaps	3				
QY	879	TATTGGTGTGATGTTGGGATCGACGAGTGTCCGCCCGGATTTTGTATCTCAACGGATC	938	1		1		1					
Db	125	TCTGGGTGGGAGCTGGTGAACAGGAGAGTGTCCGTGTAGCTGTGGTGAGCAAGAGGGGCT	184	1		1		1					
QY	939	TCGTGTATCCATCGCCACAGAAATAATCATCGACTACGGCGGCGCAGCGGAAAGCGCGGTGA	998	1		1		1					
Db	185	CCTGTGGCTTTTGTGACAGACACCATTTAAGAAGTGGGAGCCCACTTACACCATCATTA	244	1		1		1					
QY	999	ACAGTCCACCCAGGAGATCTGGCAGCGCGTGTTCATGTATTCGTATTCGATCCCTCACTCT	1058	1		1		1					
Db	245	GCAGTCTCCGAGGACATCTGGCGTGGCGTGTGTGTGCACAAAGAAAGT-----TGT	298	1		1		1					
QY	1059	GGCAGACGTTTGTGCAAAAGTGTGGCAGGCACTGGTTTTATGCCACTGTTCCTGTGT	1118	1		1		1					
Db	299	ACAAGGAGATGGATTATTAACCAAAATTCGAGGACATTGGGTGTGATGCGACATGTTCTGTGT	358	1		1		1					
QY	1119	GGTACTGATTAATAAACGGGATGCATATTCGCTGTACAGCCCGGGAAGGAGATGACAAAGCAAA	1178	1		1		1					
Db	359	TGTTTTGGATTAAGCAGTTTCACCCATTACCAAGTCAACAGAAAGGGGATTTCCATCGAAA	418	1		1		1					
QY	1179	TATCATTTGTGTGATGATGATCACCGCGGCCACCGAAGCAGAGACGATTAATGCCACTCA	1238	1		1		1					
Db	419	CGTCATCATGTTGGCTGGGACCATCGAGACATCGATCAAGTTAACAGATCAATGAGACCAA	478	1		1		1					
QY	1239	CCATCCGGTGTCTGAACACTACGTGCTGTGTAA--ATTTCGCTGTAATGAAACCGCAAA	1297	1		1		1					
Db	479	GCACAGATGCTCCCTCCAGTACGTGGGGGGGTGATGTCTGTGAATAATGACAGGCCGCAAC	538	1		1		1					
QY	1298	TTCCTCGGTGGAAGAAATATGCCAGAT---CTACACAGCTGCCGGACATTTTGTG	1354	1		1		1					
Db	539	TTCCTGTGGCTGAAGAGAACTTGAAGAGATTTGCTGGGATTAAGCGGGACATTTCTTTTG	598	1		1		1					
QY	1355	ATCTGGCCCATTTTCTGACCTGGCGGGCTACCGGATTTATTCGCGGTTAGATGACACTG	1414	1		1		1					
Db	599	ATCTCCAGACTCTTATATGTGTGGAAGGCAACAGGTGTCAACACAGCGTCTCTGCTCC	658	1		1		1					
QY	1415	TTACCTGTAAATGACGT	1432	1		1		1					
Db	659	TGCTGTGTAAATGACAT	676	1		1		1					
RESULT 8													
LOCUS	AZ302589/c	535 bp	DNA	linear	GSS 06-MAR-2001								
DEFINITION	GSSBnu1525 Brucella abortus random genomic library Brucella												
ACCESSION	AZ302589												
VERSION	AZ302589.1												
KEYWORDS	GSS.												
SOURCE	Brucella melitensis biovar Abortus.												
ORGANISM	Brucella melitensis biovar Abortus.												
REFERENCE	1 (bases 1 to 535)												
AUTHORS	Sanchez,P.,O., Zandoueni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,												
	Fachio,P., Diaz,G., Lanzavecchia,S., Aguerro,F., Frasch,A.C.C.,												
	Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.												
	Gene discovery through genomic sequencing of Brucella abortus												
	Infect. Immun. 69 (2), 865-868 (2001)												
TITLE	21101034												
JOURNAL	Small Genomes Sequencing Group												
MEDLINE	Department of Molecular Evolution, Uppsala University												
COMMENT	Notbyvagen 18C, S-752 36, Uppsala, Sweden												
	Tel: 46-18-471-4379												
	Fax: 46-18-471-6404												
	Email: Siv.Andersson@ebc.uu.se												
	Sequences were basecalled with phred and vector was masked with												
	crossmatch (see http://genome.washington.edu). Sequences were then												
	trimmed from both ends to provide low quality bases and masked												
	vector.												

FEATURES	Class: shotgun
SOURCE	Location/Qualifiers
	1..535 /organism="Brucella melitensis biovar Abortus" /strain="2308" /db_xref="taxon:235" /clone="UJI525" /clone_id="Brucella abortus random genomic library" /note="Vector: modified M13"
BASE COUNT	123 a 174 c 144 g 94 t
ORIGIN	-
Query Match	3.9%; Score 153.8; DB 17; Length 535;
Best Local Similarity	57.1%; Pred No. 3; 8e-34;
Matches 300; Conservative	0; Mismatches 222; Indels 3; Gaps 1;
OY 1716	CATGCGCATCTACCACTTCTCCCTCGTTTGATCCGGGTGTCTGCGGGACCGCTATTACAGTGC 1775
DB 533	CATGACCAACACTGAACAACGCCGTGTTTCGTGATGATGATGTCTTGCGGGCCCTATTTCGCGC 476
OY 1776	GATGCTCCAGGTCTGTGTGTTAGTTGAAGCGCGGCAAAAGTCTCGCGGAGCAGCTATTGA 1835
DB 475	CATGTGCGCGGGCTGTGTGTTGAAGAAGCGCGCAACATGCTGCGGGGCGCGTAATTGA 416
OY 1836	CCAGCTACTGTGATTTCCATCCCGGCTGTTGAABAACCTCCGGAATGCGCACACAGCTGTGA 1895
DB 415	TGATCTGATTCACATGACATCCCTTCGCGCGGGAAGCCGAAAAGGACGCGCGATCAAG 356
OY 1896	TCAGCCCCCTCCCCTGCTGCTGATGATGATCTCGAAAAACGGCGCAACATCAGA 1955
DB 355	CAAAGGGGCTGGCAGACAGCCTCGCTGCCAA---GTTGAAGCTAAGAGTGTCGAAAA 299
OY 1956	TGCTGTGCGCCCTGGCGAAAGGGGCTACAGCTGTGCGCGGAATTTCTCGAAATCGCGGCC 2015
DB 298	AACGGGCATGATCGTTGGGACATTCATGTCTGCGCGGAATTCCTCGTAATCTGCAGACC 239
OY 2016	CTTCCGACATCTCTCATGCGACAGCGGTAATTTGTGGCCCTGGGTATGAGACGAGATCTGA 2075
DB 238	TTTTCCCGCTGATGCGCGCGGCTTAATTCGCGGGGCTGGAGCCTTGATACAGGATGA 179
OY 2076	TAAATTACTGCGCTGTATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2135
DB 178	CAGTCTGGCGGACACTTTATCTGCGCGGCTTTGCGGGCTTGCTATGATGATGATGATGATGAT 119
OY 2136	TCTGACAGCTCAACAGCGCGAGGAGTAGTAGTAATAATATCTGTAATGAGCGGGGTGC 2195
DB 218	TATCGAAGGCGAGCGGCGCAAGGATATGTGTCTCACACTATTTGCTCTCGCGCGCGC 59
OY 2196	CGGCGACGATCCACTGCTGATGACGACAAATTTCTGGCAATACACTTCGG 2240
DB 58	GCGCGGCTCAATCTGCTGCGGCGAGGTCTCGCCGATGGAGCGG 14
RESULT 9	
LOCUS	AZ650403 655 bp DNA linear GSS 14-DEC-2000
DEFINITION	1M0520L01R Mouse 10kb plasmid UUC1M library Mus musculus genomic clone UUC1M0520L01 R, DNA sequence.
ACCESSION	AZ650403
VERSION	AZ650403.1 GI:11784856
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 655)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0520 row: L column: 01
Seq primer: CACACAGAAACAGCTATACCC
Class: plasmid ends
High quality sequence stop: 655.

FEATURES

source

1. 655
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0520L01"
/clone_1lb="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g147321419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 129 a 232 c 178 g 116 t
ORIGIN

Query Match

Best Local Similarity 3.8%; Score 150; DB 17; Length 655;
Matches 353; Conservative 0; Mismatches 295; Indels 6; Gaps 2;

OY 1024 GCGGCTGTTCATGATGATGCGGCTCCTGCGACAGCGTTGTGCACAAAGTGTG 1083
|||||
Db 655 GCGGCTGTTCATGCGGTCAGAGCGGCTTGCGGCGGCTGTCGCGGAAGCAT 596
OY 1084 GCGGATCGGTTTATGACCACTGCTCTGTGTGTACTGATATAAAGGTGATCCA 1143
|||||
Db 595 GCGGATCGGTTTATGACCACTGCTCTGTGTGTACTGATATAAAGGTGATCCA 536
OY 1144 TTGCGTCTGAGCGGGAAGAGATGCAAAATATCATTTGTGATGATGATCAACCG 1203
|||||
Db 535 GTTTCGCTCCGTCACCTGCAAGAACAGTGTGACACGATCTGTGCTTGACACCGC 476
OY 1204 GCCACGCAACAGCAGAGCAATCAATGCCATCACCATCGGTGCTGAACACTACGCGT 1263
|||||
Db 475 GCATGCTGAGGCGCATGCTGACGCGATCGGCGACCGGGTTTGGATTTTGGCGGT 416
OY 1264 GGTAAATTTTGGCTAAATGGAACCGAAATTTCTGTGCTGAAGAAATATGCCA 1323
|||||
Db 415 AACAGCGTGTGCGCGGAATATGAGATGCGAATGATGCGTGAAGAAATTTGCGC 356
OY 1324 GAGATCTAGAACGTCGCCGGAATTTTTCGATCTGCGCATTTTTCGATCTGCGCGGT 1383
|||||
Db 355 GCAAGCTGGGCGCGCATGTCTTTCGCTTCGATCTGCGCATTTTCTCACCTTGAAAGCG 296
OY 1384 ACCGCTGATTTACGCGCTTCAGTATGACCTGTAAATGAGACGTGGCGACAT 1443

Db 295 ACCGCTTCGCGCAAGCGTTCAATTCACGACGAGCGCAAGTGAATTTCTCGCGCAG 236
OY 1444 GAAATTCGCTGGGATCCAGATTTATTTCCGACCATTTGCCCTTCACAGTATGAGTAA 1503
|||||
Db 235 GAAATTCGCGGTTGGCAGGCGGATTTATCTC-GAAGTGGCGGCGGCTGTGATCTGAAGG 177
OY 1504 GATTTATTCGCAATGTCATCATTTATTTTCCTCCGGAACACCTGTTGAAATGTTTA 1563
Db 176 ACAGGCGCGGCTGCGGACACCGATGATGCGGCGGGAAGCATGCTCCGTG-121
OY 1564 ACAGCAACGCCGCGGACAGATGGATTAATCTCCGCGCACACCTGTGCTGATGAGTTA 1623
Db 120 -CGCTGAAGCGGCGGACAGAACTTGCTGATACGAGGTTGCCAGTGGCGGCGGCGATG 62
OY 1624 ATGATGCTTCACGCTGTGTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTG 1677
Db 61 ATGATGCTTCACGCTGTGTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 8

RESULT 10

BI557944 712 bp mRNA linear EST 05-SEP-2001
LOCUS 603237561F1 NC1_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5290478 5',
DEFINITION mRNA sequence.

ACCESSION BI557944
VERSION BI557944.1 GI:15445258
KEYWORDS EST.

SOURCE

house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 712)
NIH-MGC http://mgi.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

COMMENT

Tissue Procurement: Lotmar Hemilghausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://imgc.lnl.gov
plate: LLMH1734 row: 3 column: 15
High quality sequence stop: 710.

FEATURES

source

1. 712
Location/Qualifiers
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5290478"
/clone_1lb="NC1_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site:1: Salt; Site:2: Mott; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lotmar Hemilghausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

BASE COUNT

174 a 152 c 211 g 175 t

Query Match

Best Local Similarity 3.8%; Score 150; DB 13; Length 712;
Matches 323; Conservative 0; Mismatches 240; Indels 9; Gaps 2;

OY 879 TATTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 938
|||||
Db 100 TGTGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 159

QY 939 TGTGCTATCCATGATCCACAGAAAAATACAGACTACGGCGCAGCCGAGCCGCTGGA 998
160 CATTGTTGGCTTTGGCAGACAGCCCAATTAAAGTGGAGCCTCAATTCAACCCACGCA 219
QY 999 ACAGTCCAGCAGAGATCTGGCAGCGGCTGTCTTATCTATTCGTATATCGCTACTCT 1058
220 GCAGTCTCTAGAGGATATCTGGCGGCGCATGCTGCC-----TTGTCAAAAGGAAGTTGT 273
QY 1059 GAGAGACGTTGTGTGACAAAGTGTGGCAGGCATCGGTTTGTATGCGCCACTGTTCTGTGT 1118
274 TCAAGGATTTGATGACATGATCGAATCCAGACCTGTTTGTATGCCACGTTGTCTGTGT 333
QY 1119 GGTACTGTGATTAAGACGCTGATCCATTTGCTGTACGCCCGGAAGAGATCAAGCAAAA 1178
334 TGTCTTGATTAAGAGATTCATCTCTTACAGATAAACCATGAAGGGAGATCTCTGTGAAA 393
QY 1179 TATCATTTGTGTGATGATACCGCGCCACCGAAGCAGAGCATCAATGCTCACTCA 1238
394 CGTTATCATGTGTGCTGACACCGGCTGTACACGAGTCCACAGATTAATGAGACCAA 453
QY 1239 CCATCCGCTGCTGACACTACGCTGCTGTAATTTGCTGTAATGGAACACCGAAT 1298
454 GCACAGAGTCTTCTGATGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 513
QY 1299 TCTCTGCTGTAAGAAATATGCGCAGAGAT--CTACGAACGTTGCCGACATTTTTCGA 1355
514 CTTGTGCTGTAAGAGAACTTGAAGAGATTTGCTGCTGATTAAGCGGACATTTCTTGA 573
QY 1356 TCTGGCGGATTTTCTGACCTGCGGGGCTACCGGTGATTTAGCGGCTTCAATGACACTGT 1415
574 TCTCCAGAGACTCTTATCATGATGAAGCAACAGAGTGTCTACAGAGTCTCTGTCTCTT 633
QY 1416 TACCTGTAAATGACGCTGCTGCGACATGA 1447
634 GGTGTCAATGACATACACAGACAGCA 665
Db

RESULT 11
B1149014 714 bp mRNA 11linear EST 05-JUL-2001
LOCUS B1149014
DEFINITION 602911220F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5052317 5',
mRNA sequence.
ACCESSION B1149014
VERSION B1149014.1 GI:14609015
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 714)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M1142 row: 9 column: 06
High quality sequence start: 7
High quality sequence stop: 702.
Location/Qualifiers
1..714
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5052317"
/clone_11b="NCI_CGAP_L19"

FEATURES
Source

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 175 a 153 c 219 g 167 t
ORIGIN
Query Match 3.8%; Score 149.2; DB 13; Length 714;
Best Local Similarity 56.2%; Pred. No. 1,1e-32;
Matches 324; Conservative 0; Mismatches 243; Indels 9; Gaps 2;
QY 879 TATTGTTGTTGATGTGGATCAGCAGCTGTCCGCCGCGGATTTTGTATCTCAACGATC 938
121 TGTGGGATGATGTGGAGACAGTACTGTCCGCGGCTGTAGTGAACAGAGAGGCT 180
QY 939 TCTGTATCCCATGCCCACAAAAATACAGACTACGCGCGCAGCGAGCCGCTGGA 998
181 CTTTGTGGCTTTGGCAGAGCGCAATTAAGAGTGGAGCCTCAATTCAACACGACGA 240
QY 999 ACAGTCCAGCAGAGATCTGGCAGGCGCTGTCTGTATCTATTCGTATGCTCACTCT 1058
241 GCAGTCTCAGAGATATCTGGCGGCTATGCTGCC-----TTGTCAAAAGGAAGTTGT 294
QY 1059 GGCAGACGTTTGTGCACAAAGTGTGGCAGGCTGCTTTGTATGACACCTGTTCTGTGT 1118
295 TCAAGGATTTGATGACATCGAATCCGAGCATTTGTTTGTATGACACGTTCTCTGTGT 354
QY 1119 GGTACTGTGATTAAGACGCTGATCCATTTGCTGTACGCCCGGAAGAGATCAAGCAAAA 1178
355 TGTCTTGATTAAGAGTTCATCTCTTACAGTAACCAATGAAGGAGATCTCTGTGAAA 414
QY 1179 TATCATTTGTGTGATGATACCGCGCCACCGAAGCAGAGCATCAATGCTCACTCA 1238
415 CGTTATCATGTGTGCTGACACCGGCTGTACACGAGTCCACAGATTAATGAGACCAA 474
QY 1239 CCATCCGCTGCTGACACTACGCTGCTGTAATTTGCTGTAATGGAACACCGAAT 1298
475 GCACAGAGTCTCTGATGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 534
QY 1299 TCTGTGCTGTAAGAAATATGCGCAGAGAT--CTACGAACGTTGCCGACATTTTTCGA 1355
535 CTTGTGCTGTAAGAGAACTTGAAGAGATTTGCTGCTGATTAAGCGGACATTTCTTGA 594
QY 1356 TCTGGCGGATTTTCTGACCTGCGGGCTACCGGTGATTTAGCGGCTTCAATGACACTGT 1415
595 TCTCCAGAGACTCTTATCATGATGAAGCAACAGAGTGTCTACAGAGTCTCTGTCTCTT 654
QY 1416 TACCTGTAAATGACGCTGCTGCGACATGAATCG 1451
655 GGTGTCAATGACATACACAGACAGCA 690
Db

RESULT 12
BF440790 503 bp mRNA 11linear EST 01-DEC-2000
LOCUS BF440790
DEFINITION BF440790 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF440790
VERSION BF440790.1 GI:11500703
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
1 (bases 1 to 503)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Db	357	TGTCGTGGATAAAGAGTTTCACCTCTTACCAATGAACCATGAAGGGACCTCTCGAAA	416
Qy	1179	TATCATTTGTGGATGATGATACACCGGGCCACGACAACAAGACGAGGCCATATATGCCATCTCA	1238
Db	417	CGTTATCATATGTGGCTGGAGACACCGGGGCTGTGCGCCAGATGCCACAGATTAAAGAACCA	476
Qy	1239	CCATCCGGTGTCTGAACTACGTGCTGTGATAAATTTTCGCTGTGAATGAAGAACACCCAAAT	1298
Db	477	GCACAGAGTCCCTTCAGTATGTGGGTGTGTATGTCTGTGGAGATGCAAGGCTCCAAAGCT	536
Qy	1299	TCTCTGGCTGGAAGAAATATGCCGAGAT---CTACGACACTGCCGCGACATTTTTCGA	1355
Db	537	CCTGTGGCGCTGAAGGAACTTGAGAGATTTTGTGGGATAAGGGGGACATTTCTTTGA	596
Qy	1356	TCTGGCCGATTTTTCAGACTGGCGGGCTACCGGTATTTAGGCGCTTCAGTATGACACTGT	1415
Db	597	TCTCCCAACACTCTTATCATGTGAAGGACAACAGTGTCAACGACACGGTCTCTCTGCTCTT	656
Qy	1416	TACCTGTAAATGGACGT	1432
Db	657	GGTGTGCAAATGGACAT	673

RESULT 14	LOCUS	DEFINITION	ACCESION
BM646051	BM646051	17000687317822 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone 19600449631795 5', mRNA sequence.	670 bp mRNA linear EST 26-FEB-2002 U945602

Accession	BM646051.1
Version	GI:18945562
Keywords	EST.
Source	African malaria mosquito.
Organism	Anopheles gambiae

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 670)	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab			
	'R., Collins, F.H., Venter, J.C. and Hoffman, S.L.			
		Celera Anopheles gambiae EST project		
		Unpublished (2002)		
	Contact: Holt R.A.			

Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel.: 2404533151
Fax: 2404534580
Email: holtra@celera.com
Plate: N0010049M1 row: p column: 01
Seq primer: M13 Reverse.

```

FEATURES
source
location/Qualifiers
1..670
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449631795"
/clone_lib="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
```

BASE COUNT	146 a	184 c	212 g	128 t
ORIGIN				
Query Match		3.6%;	Score 144..4;	DB 13; Length 670;
Best Local Similarity		56.0%;	Pred. No. 2..9e-31;	
Matches 337; Conservative		0;	Mismatches 256;	Indels 9; Gaps 3;

Qy 1086 AGCATCGTTTGTATGCCACCTGTTCTCTGGTGGTAC^uGGATAA^uAAACGGTGATCCATT 1145

Db 32 AGGCATCGGGTTCGATGCGACCGTTCACCTGTTGTTCTCGATGGCCAGATGCACCGCT 91

QY 1146 GCCTGTCAGCCCGAGAGAGATGCCAAGCAA---AATATCATTTGTGTGATGATCACC 12

Db 92 AACGGTCTGCCCGTCTTGAAGGACAACAGCCACGATGATCCTCTGATGGACCATCG 15

1203 GGCACCCGAACACAGAGAGCGGAGCAGATGCACATCCACATCCGGTGTCTGACTACGTCGG 12

QY 1263 TGGTAAATTTCCGCTGAATGGAACACCGAAATCTCTGGCTGAACGAATATGCC 13

Db 212 TGGCAGCATTTCCGCTCGAGATGCAAGTGGCCGAAGCTGCTGTGGCTCAAGCGAACAATGCA 277

1323 AGA---GATCTACGACGTCGCCGACATTTTTCGATCTGCGCGATTTCCTGACCTGGCG 13

1 380 GGCTACCGGCGATTTAGCGCGCTTCAGTATGCACTGTATACCTCGTAATGAGACCTGGC 14

Db 332 TGGCACTGGTGGCCGATTGGCCGATTCGTCGCAATGGACTACGATGC 39

QY 1440 ACATGAAATCGCTGGGATCCAGATTATTTCGGCACCATTGG---CCTTGCAGAGTTAGC 14

Db 392 AGAGCATCGATGCTGGATTTCTCTGCAACGATCGGTTTGGTCGATGATCTACG 45

QY 1497 GGAGGAGAGATTTTATTCGCATGTCATCATCATTTGTTTCTCCCGGACACACTTGTGAAA 15

QY 1557 TGGTTTACAGCACAAGCCGGCGCAGAGATGCGATTACTCCCGCGGCACACCTGTGCGTGT 16

Db 512 AGGACTTCGGCGCAAGCGGCCAAGAGCTGGGGTTAAACCTTGCGACGGCAGTAGCAAG 57

[illegible]

Db 572 TTCGATGATCGATGCCGATGCGGAGCGTTGGCACTGCTTGATGCGAAGCACCACCGGG 63

xy	10/7	GA	10/8
Db	632	GA	633

LOCUS	BP	mRNA	linear	EST
BQ932942	983			21-AU
RESULT 15				

DEFINITION	AGENCOCURT_8782772 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6357, mRNA sequence.
------------	--

ACCESSION	B0932942
VERSION	B0932942.1
KEYWORDS	GI:22348325
FAST	

ORGANISM	human.
SOURCE	Homo sapiens

REFERENCE
1 (bases 1 to 983)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

ALL-INDIAN	1
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csrausberg@aol.com

Email: cyapus-1@mail.ru, hinh.gov
 Tissue Procurement: jcdtd@rnp.gazdar
 CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLCM2553 row: h column: 11
High quality sequence stop: 530.

```
FEATURES      Location/Qualifiers
source        1..983
```


GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:37 : Search time 20.4884 Seconds
(without alignments) 1625.927 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MMHNSVPSMNTPLNGKVAAL.....TRSKNVTRDVLVPGSVDL 250

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq_101002:*

- 1: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	100.0	250	22	AAE09779
2	366	28.9	247	16	AAE07866
3	366	28.9	248	22	AAE07914
4	358	28.3	240	16	AAE07914
5	342	27.0	251	21	AAV55986
6	319.5	25.3	231	22	ABP39552
7	311	24.6	337	22	ABP39552
8	311	24.6	337	22	AAU23036
9	311	24.6	337	22	AAU18432
10	311	24.6	337	22	AAU42410

11	309	24.4	310	21	AAE27650	Human protein PRO2
12	309	24.4	310	22	AAE80238	Human PRO238 prote
13	309	24.4	310	23	ABE95441	Human angiotensin
14	309	24.4	310	23	ABE84835	Human PRO238 prote
15	309	24.4	325	21	AAV92513	Human OXRE-10. Ho
16	309	24.4	325	21	AAV73426	Human secreted pro
17	309	24.4	325	22	AAE38738	Human polypeptide
18	309	24.4	327	22	AAU18303	Human polypeptide
19	309	24.4	327	22	AAE40524	Human endocrine po
20	309	24.4	350	19	AAW74861	Human polypeptide
21	303	24.0	310	20	AAV13370	Human secreted pro
22	292	23.1	248	23	ABE47755	Human acid sequenc
23	291.5	23.0	254	22	AAE47522	Listeria monocytog
24	285	22.5	247	23	ABE48992	(R)-2-octanol dehy
25	285	22.5	248	23	AAV54422	Listeria monocytog
26	281.5	22.3	269	23	ABE92448	Amino acid sequenc
27	281.5	22.3	280	23	ABE92449	Herbicidally activ
28	278	22.0	206	23	AAU99345	Short-chain dehydr
29	275.5	21.8	253	23	ABE54214	Lactococcus lactis
30	275.5	21.8	272	21	AAV54420	Secoisolaricresin
31	275.5	21.8	273	21	AAV54412	Secoisolaricresin
32	273	21.6	202	23	AAU77210	Consensus protein
33	272.5	21.5	277	21	AAV54415	Secoisolaricresin
34	271	21.4	260	21	AAE51581	Arabidopsis thalia
35	271	21.4	260	23	ABE92475	Arabidopsis thalia
36	271	21.4	261	21	AAE24387	Arabidopsis thalia
37	271	21.4	274	21	AAE51580	Arabidopsis thalia
38	271	21.4	276	21	AAE24386	Arabidopsis thalia
39	271	21.4	286	21	AAE51575	Arabidopsis thalia
40	270.5	21.4	251	22	AAU34193	Staphylococcus aur
41	270.5	21.4	272	22	AAU37095	Staphylococcus aur
42	267.5	21.1	277	21	AAV54413	Secoisolaricresin
43	265	20.9	250	23	ABE54881	Lactococcus lactis
44	262.5	20.8	241	22	ABE96397	Putative P. abyssi
45	261	20.6	249	23	ABP39980	Staphylococcus epi

ALIGNMENTS

RESULT 1	AAE09779	standard; Protein; 250 AA.
ID	AAE09779	
XX	XX	
AC	AAE09779;	
XX	XX	
DT	29-NOV-2001	(first entry)
XX	XX	
DE	Escherichia coli strain C rfl operon encoding ribitol dehydrogenase.	
XX	XX	
KW	Positive selection system: metabolise; arabitol; ribitol; mannitol;	
KW	transgenic cell; marker gene; ribitol dehydrogenase; ribitol kinase;	
KW	ribitol transporter; rfl operon.	
XX	XX	
OS	Escherichia coli C.	
XX	XX	
PN	W0200166779-A2.	
PD	13-SEP-2001.	
XX	XX	
PF	08-MAR-2001; 2001WO-US07474.	
XX	XX	
PR	08-MAR-2000; 2000US-0188291.	
XX	XX	
PI	15-AUG-2000; 2000US-0255595.	
XX	XX	
XX	(UYGE-) UNIV GEORGIA RES FOUND INC.	
XX	Parrott W, Lafayette P, Kane P;	
DR	WPI: 2001-565596/63.	
XX	N-PSDB; AAD16811.	
PT	Positively selecting transformed cells comprising selectable marker	

PT gene and desired gene, from a cell population by using marker compounds
PT e.g., arabinol, ribitol which confer selective advantage on transformed
XX cells
PS Claim 16; Page 37; 37pp; English.
XX
CC The present invention relates to a positive selection system that
CC involves conferring to transferred cells the ability to metabolise
CC arabinol, ribitol and/or mannitol. The positive selection method is
CC used in positively selecting transgenic cells from a population of cells
CC using the positive selection method, the presence of the gene of
CC interest in the genetically transformed cells may be determined without
CC the disadvantages associated with traditional negative selection
CC systems. Positive selection of the transformed cells is achieved without
CC directly damaging the neighbouring non-transformed cells. The
CC transformed cells may be identified by simple visual means without the
CC use of a separate assay to determine the presence of a marker gene. This
CC technique also avoids the release of antibiotics or other dangerous
CC genes into the environment. The present sequence is Escherichia coli
CC strain C ribitol dehydrogenase encoded by ribitol operon (rbl operon).
CC The operon also encodes ribitol kinase and ribitol transporter.
XX
SQ Sequence 250 AA;

Query Match 100.0%; Score 1265; DB 22; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.4e-127;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMNHSVSMPTPLNGKVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIG 60
Db 1 MMNHSVSMPTPLNGKVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIG 60
QY 61 ENAVALQLDLFNNQVDMNLADIETELAGLDIFHANAGAYIGGVAEGDPDVMRVLNLN 120
Db 61 ENAVALQLDLFNNQVDMNLADIETELAGLDIFHANAGAYIGGVAEGDPDVMRVLNLN 120
QY 121 INAAFRVAVLPHMIAORSDDIETSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVS 180
Db 121 INAAFRVAVLPHMIAORSDDIETSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVS 180
QY 181 QYGVAVGAVLPGPVVATLDDMPKAKMEALANGSLMQPIEVAESVLFMTYTRSKNVTVRD 240
Db 181 QYGVAVGAVLPGPVVATLDDMPKAKMEALANGSLMQPIEVAESVLFMTYTRSKNVTVRD 240
QY 241 LVILPGSVDL 250
Db 241 LVILPGSVDL 250

RESULT 2
AA077866
ID AAR77866 standard; Protein; 247 AA.
XX
AC AAR77866;
XX
DT 13-NOV-1995 (first entry)
XX
DE S. clavuligerus ORF9 product.
XX
KM Clavulanic acid; clavulinate; antibiotic; beta-lactamase-inhibitor.
XX
OS Streptomyces clavuligerus.
XX
PN CA2108113-A.
XX
PD 09-APR-1995.
XX
PF 08-OCT-1993; 93CA-2108113.
XX
PR 08-OCT-1993; 93CA-2108113.
XX
PA (UYAL-) UNIV ALBERTA.
XX

PI Aideo KA, Jensen SE, Parakkar AS;
XX
DR WPI: 1995-207301/28.
DR N-PSDB: AA091580.
XX
XX
PT Clavulanic acid biosynthesis enzymes and corresp. DNA - useful for
PT biosynthesis of the antibiotic in Streptomyces hosts which do not
PT naturally produce clavulinate
PS
PS Claim 31; Fig.18; 41pp; English.
XX
CC A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (AA091580),
CC extending downstream from pcbc, included 10 ORFs encoding the
CC enzymes required for clavulinate biosynthesis. ORF8 (oriented
CC in the opposite direction to pcbc) encoded a protein (AAR77866)
CC that showed weak similarity to ribitol 5 P04 dehydrogenase-type
CC enzymes.
XX
SQ Sequence 247 AA;

Query Match 28.9%; Score 366; DB 16; Length 247;
Best Local Similarity 37.3%; Pred. No. 7e-31;
Matches 91; Conservative 41; Mismatches 104; Indels 8; Gaps 3;

QY 9 MNTPLNKRVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEI--GENAYA 65
Db 1 MNTPLNKRVAATGAAGTGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEI--GENAYA 65
QY 66 LQDLFNNQVDMNLADIETELAGLDIFHANAGAYIGGVAEGDPDVMRVLNLNAAAF 125
Db 61 LELDVADRGVDAVAATVEALGSLDLVNNAGTLMGLPVEDADTPTWTMTIDTNLLGLM 120
QY 126 RCYAVLPHMIAORSDDIETSSIAGVVPVIMEPIYTASKFAVOAFVHTTRQVSQYGV 185
Db 121 YMTAALPHILRSK-GTVVQMSIAGVAVNNAVYQATKFGVNAFSETLRQEVTERGVR 179
QY 186 VGAVLPGPVVATL---LDDMPKAKMEALANGSLMQPIEVAESVLFMTYTRSKNVTVRD 241
Db 180 VVYIEPPTTTELGHITHTATKEMVDRISQIRKQADIAEAVRYAVTAPRHATVHEI 239
QY 242 VILP 245
Db 240 FIRP 243

RESULT 3
AA07914
ID AAE07914 standard; Protein; 248 AA.
XX
AC AAE07914;
XX
DT 01-NOV-2001 (first entry)
XX
DE S. clavuligerus clavulanic acid biosynthesis enzyme #6.
XX
XX Clavulanic acid biosynthesis enzyme; antibiotic; infectious disease;
XX broad spectrum beta-lactamase inhibitor; open reading frame; ORF;
XX pcbc gene.
XX
OS Streptomyces clavuligerus.
XX
FH Key Location/Qualifiers
FT Misc-difference 247..248
FT /note="Encoded by GTC on the inverse complementary
XX strand of the corresponding DNAs (AAD14499, AAD14510)"
XX
PN US6232106-B1.
XX
PD 15-MAY-2001.
XX
PF 30-AUG-1999; 99US-0385028.
XX
PR 29-JAN-1997; 97US-0790462.
XX

PR 08-OCT-1993: 93US-0134018.
PR 06-DEC-1995: 95US-0567801.
XX
PA (UYAL-) UNIV ALBERTA.
XX
XX
PI Jensen SE, Aldoo KA, Paradkar AS;
XX
DR WPI, 2001-342772/36.
XX N-PSDB; AAD14499, AAD14510.
DR
XX
PT Novel enzyme required for clavulanic acid biosynthesis which is useful
PT as broad spectrum beta-lactamase inhibitor -
XX
XX
PS Claim 1; Fig 17: 75pp; English.
XX
XX The invention relates to DNA sequences encoding enzymes required
CC for clavulanic acid biosynthesis. Clavulanic acid is a broad spectrum
CC beta-lactamase inhibitor and is an important antibiotic for the treatme
CC of infectious diseases. Also provided in the patent is a 15 Kb genomic
CC DNA fragment downstream to pcbc gene from streptomyces clavuligerus.
CC The genomic DNA comprises 10 open reading frames (ORFs), eight of which
CC are involved in clavulanic acid biosynthesis. The present sequence is
CC S. clavuligerus clavulanic acid biosynthesis enzyme encoded by ORF.
XX
XX Sequence 248 AA;

	Query Match	28.9%	Score 366;	DB 22;	length 248;
	Best Local Similarity	37.3%;	Pred. No. 7e-31;		
	Matches	91;	Conservative	41;	Mismatches 104; Indels 8; Gaps 3.
QY	9	MNTPLNGVVAATITGAASGIGLOCAKTLTDAGAKVVLIDREGDKLHKIYAEI--GENAYA	65		
		:			
Db	1	MSAALGGKVALITITGASSGIGENATAPALAAEGAAVVAIAARVRVKYLAIGDELTAACAGAKYAV	60		
		:			
QY	66	LQLDLFNNQOVDMADIIETLAGGDIIFHANGAYITGCGVAGGDDPVMQDRLVNLNTNAAF	125		
		: : : : : : : : : : : : : : : : : : : :			
Db	61	LELDVADQGVDAANASTVEALGGDILVNNNGIMLGGVEADPTDWTIRMIDTNLGLM	120		
		: : : : : : : : : : : : : : : : : : : :			
QY	126	KCVRAVLPHMTAQRSGDIIFTSSINGVVPVITPEYITASKFAVQAFVHTTRQVSOGYVR	185		
		: : : : : : : : : : : : : : : : : : :			
Db	121	YMTAALRPHLLNSK-GTVVQMSISINGRVNVRNAAVYQAIRKFCVNAFSETLRQEVTERGYVR	179		
		: : : : : : : : : : : : : : : : : : :			
QY	186	VGAVVLPGEVNTAL---LDDMPKAKMEKALANGSLMOPTEVAESVLEWYTRSKNTVARDL	241		
		: : : : : : : : : : : : : :			
Db	180	VVIVIEPGTDTDFRLRHIIHTATKKEVFEQRISQIRKLQADIDAEVRYATATAPHHATVHEI	239		
		:			
QY	242	VILP 245			
Db	240	FIRP 243			

	RESULT 4
AAR61477	ID AAR61477 standard; Protein; 240 AA.
XX	
XX	AAR61477;
AC	
XX	
DT	01-OCT-1995 (first entry)
XX	
DE	Clavulanic acid dehydrogenase sequence
XX	
KW	Clavulanic acid; antibiotic; Augmentin
XX	
OS	Streptomyces clavuligerus ATCC 27064.
XX	
PN	M09503416-A.
XX	
PD	02-FEB-1995.
XX	
PE	15-JUL-1994; 94MO-EP02346.
XX	
PR	24-JUL-1993; 93GB-0015393.
XX	

PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Arnell J, Elson SW, Nicholson NH, Woroniecki SR;
PI
XX WPI: 1995-075242/10.
DR
XX N-PSDB: AAO67531. |
DR
XX
PT New clavulanic acid dehydrogenase from Streptomyces, and related
PT DNA and vectors - used to produce beta-lactamase inhibiting
PT clavulanic acid from new 3-oxo:ethylidene analogues
XX
XX
PS Claim 2; Fig 1; 36pp: English.
XX
XX A new enzyme is disclosed which has clavulanic acid dehydrogenase
CC activity, the enzyme having an apparent mol. wt. of 28 kD (by SDS
CC PAGE) and incorporating the N-terminal sequence PSLDGGKVALRIGASSIGE.
CC The enzyme is derived from the mycelium of a Streptomyces species,
CC especially *S. clavuligerus* (e.g. ATCC 27064), *S. jumojensis* (e.g. ATCC
CC 29864) or *S. katsurahamanus* (e.g. T722). The present sequence
CC represents the enzyme from *S. clavuligerus* ATCC 27064.
CC The enzyme is used to catalyse the biosynthesis of clavulanic acid
CC from a precursor clavulanic acid aldehyde. The obtained clavulanic acid
CC is in turn a key ingredient in the antibiotic Augmentin.
XX
XX Sequence 240 AA;

[illegible]

RESULT	5
AAV55986	
ID	AAV55986 standard; Protein; 251 AA.
XX	
AC	AAV55986;
XX	
DT	15-MAR-2000 (first entry)
XX	
DE	S.clavuligerus cladh protein.
XX	
KW	Operon; cladh; clar; biosynthetic pathway; clavulanic acid; enzyme;
RW	clavulanate-9-aldehyde reductase; transcriptional regulator;
KW	gene expression.
XX	
OS	Streptomyces clavuligerus.
XX	
PN	ES213I001-A1.
XX	
PD	01-JUL-1999.
XX	
PE	16-JUN-1997; 97ES-0001305.
XX	
ER	16-JUN-1997; 97ES-0001305.
XX	
PA	(ANTI) ANTIBIOTICOS SAU.
XX	

DR	WP1: 2000-026016/03.	
DR	N-PSDB: AAZ30700.	
XX		
PT	Improving production of clavulanic acid by Streptomyces - by	
PT	super-expression of clar gene.	
XX		
PS	Example 1: Page 8-9; 23pp; Spanish.	
XX		
CC	This sequence represents the cladh protein which is a putative	
CC	clavaninate-9-aldohdeyde reductase used in the biosynthetic pathway for	
CC	clavanilic acid in Streptomyces clavuligerus. The invention relates to	
CC	methods of improving production of clavulanic acid by Streptomyces by	
CC	super-expression of clar gene. The clar gene is characterised in that	
CC	it is localised in the gene group encoding genes for biosynthesis of	
CC	clavanilic acid.	
SO	Sequence 251 AA:	
	Query Match 27.0%; Score 342; DB 21; Length 251;	
	Best Local Similarity 35.7%; Pred. No. 2.7e-28;	
	Matches 87; Conservative 40; Mismatches 109; Indels 8; Gaps 3	
OY	9 MMTPLNGKVAATITGAASGIGLOCATTLDDAGAKKVLIDREGKRLKHYAEL--GENAYA 65	
Db	5 MPSALQKVALTTGREGLGHRATRALAPEGAAVIAAARVERKRLGDELTAAGAKKAV 64	
OY	66 LQLDLENNQOVNMMADIIELAGDIFPANGAVATGCGVAGCDPPVMDRVNLNTNMAF 125	
Db	65 LELDVADNQGVAAVAASYAEALGGLDILVNNNGIMLGPVEADDTTDMTRMIDTNLGLM 124	
OY	126 RCVRAVLPHMTAQRSGDIIETSSIAGVVPIWEPIYTASKFAVQAEVHTTRQVSQYGV 185	
Db	125 YMTRAVLPHLRLSK-GTVVQMSISAGRVTVRNAAVYQATKFCVNAFSETVROEVEERGV 183	
OY	186 VGAVLPGPVVTAL---LDDMPKAMEEALANGSLMOPTEVAESVLFVTRBSKNVTVRDL 241	
Db	184 VVIVIEPGTTDTLRLGHTHTTATKMEYEQKISQIRKLQADIDAEAVRAVATAPHHATVHET 243	
OY	242 VILP 245	
Db	244 FIRP 247	
RESULT 6		
XX	ABP39552	
ID	ABP39552 standard; Protein: 231 AA.	
XX	ABP39552;	
XX		
DT	24-JUL-2002 (first entry)	
XX		
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4397:	
XX		
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;	
KW	antibacterial; gene therapy.	
XX		
OS	Staphylococcus epidermidis.	
XX		
PN	US6380370-B1.	
PD		
XX	30-APR-2002.	
PF	13-AUG-1998; 98US-0134001.	
XX		
PR	14-AUG-1997; 97US-055779P.	
XX	08-NOV-1997; 97US-064964P.	
XX		
PA	(GENO-) GENOME THERAPEUTICS CORP.	
XX		
PI	Doucette-Stamm LA, Bush D;	
XX		
XX	WPI: 2002-381255/41.	
DR	N-PSDB: ABN92097.	

xx		Novel isolated nucleic acid encoding a staphylococcus epidermidis
pt		polypeptide, useful for diagnosing and treating bacterial infections -
ps		
px		Disclosure; SEQ ID #397; 267bp; English.
cc		
cc		ABN90538 to ABN9374 represent Staphylococcus epidermidis open reading
cc		frame (ORF) nucleic acid sequences which encode the amino acid sequences
cc		given in ABP5124 to ABP37960. The S. epidermidis sequences have
CC		antibacterial activity and can be used in gene therapy. The sequences
CC		can also be used in the diagnosis and treatment of bacterial infections,
CC		particularly S. epidermidis infections. The sequences can be used to
CC		screen for compounds able to interfere with the S. epidermidis life
CC		cycle or inhibit S. epidermidis infection.
CC		N.B. The sequence data for this patent did not form part of the printed
CC		spectification, but was obtained in electronic format directly from the
CC		USPTO web site.
xx		
SQ		Sequence 231 AA;
	Query Match	25 3%; Score 319.5; DB 23; Length 231;
	Best Local Similarity	30.2% Pred NO. 6.4e-26;
Matches	70; Conservative	59; Mismatches 92; Indels 11; Gaps 2
OY		16 KVAATITGAASGIGLOCAKYLIDAGAKVVLIDDEGRKLHKIVLELGENAVAYALDLENNOQ 75 : : : : : : : : : : : : : : : DB 8 KAAYVTGASSSGICEAIANKLSQGASISYLVGNREGRLNEIYOQLNNPAKVTVADTYKS N 67 OY 76 VDNMLADIETLAGGLDIPIRANNGAIVIGEVAAGDPDDWDRVNLININAFAFCRAVALPMH 135 : : : : : : : : : : : : : : : DB 68 IDDMKAAVIDEIGHIDIYNVSAGSLSSKITDYANEOWTMDIVNKGTJHLVLAATLPYL 127 OY 136 IAORGSDIEFTSIAGVVPVIVEIPYTASKFEVAOVAFHTRPROVSOYGVRGAVLPGPVY 195 : : : : : : : : : : : : : DB 128 LKGSSHIIINLASVGSEFKTNNAVYGATKAIHNIHQGLEELARTGYKVASISPQWD 187 OY 196 TALLD-DMPKAKMEBALANGSLMOPIEVAESYLEMTRSKRNVTVRDVIYP 245 : : : : : : : : : : : : : : : DB 188 TPWTGTGFGEGERKKLEAQ-----NTADVVYVALTOPSHVNVEVIRP 230 RESULT 7 ABBI0190 ID ABBI0190 standard; Protein; 337 AA. XX AC XX ABBI0190; DT 10-JAN-2002 (first entry) XX DE Human CDNA SEQ ID NO: 498. XX Human; gene therapy; neural disorder; immune system disorder; KW pulmonary disorder; reproductive disorder; gastrointestinal disorder; KM proliferative disorder; inflammation. XX OS Homo sapiens. XX PN WO200154474-A2. XX PD 02-AUG-2001. PE 17-JAN-2001; 2001WO-US01349. PR 31-JAN-2000; 2000US-179065P. PR 04-FEB-2000; 2000US-180628P. PR 24-FEB-2000; 2000US-184664P. PR 02-MAR-2000; 2000US-186350P. PR 16-MAR-2000; 2000US-189874P. PR 17-MAR-2000; 2000US-190076P. PR 18-APR-2000; 2000US-198123P. PR 19-MAY-2000; 2000US-205515P. PR 07-JUN-2000; 2000US-209467P.

PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226681P.
PR 22-AUG-2000; 2000US-226688P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232080P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.
PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 14-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234977P.
PR 25-SEP-2000; 2000US-234988P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235835P.
PR 29-SEP-2000; 2000US-236377P.
PR 29-SEP-2000; 2000US-236378P.
PR 29-SEP-2000; 2000US-236379P.
PR 29-SEP-2000; 2000US-236380P.
PR 29-SEP-2000; 2000US-236381P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.

PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 17-NOV-2000; 2000US-248207P.
PR 17-NOV-2000; 2000US-248208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249246P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251858P.
PR 08-DEC-2000; 2000US-251859P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
(HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM.
XX
PI
XX
DR WPI; 2001-476161/51.
DR N-PSDB; ABA06412.
XX
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition
XX
XX
XX Claim 11; SEQ ID NO: 498; 859pp + Sequence Listing; English.
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.

PR 07-JUL-2000: 2000US-0216647.
 PR 07-JUL-2000: 2000US-0216880.
 PR 11-JUL-2000: 2000US-0217487.
 PR 11-JUL-2000: 2000US-0217496.
 PR 14-JUL-2000: 2000US-0218290.
 PR 26-JUL-2000: 2000US-0220963.
 PR 26-JUL-2000: 2000US-0220964.
 PR 14-AUG-2000: 2000US-0224518.
 PR 14-AUG-2000: 2000US-0224519.
 PR 14-AUG-2000: 2000US-0225213.
 PR 14-AUG-2000: 2000US-0225214.
 PR 14-AUG-2000: 2000US-0225266.
 PR 14-AUG-2000: 2000US-0225267.
 PR 14-AUG-2000: 2000US-0225268.
 PR 14-AUG-2000: 2000US-0225270.
 PR 14-AUG-2000: 2000US-0225447.
 PR 14-AUG-2000: 2000US-0225757.
 PR 14-AUG-2000: 2000US-0225758.
 PR 14-AUG-2000: 2000US-0225759.
 PR 18-AUG-2000: 2000US-0226279.
 PR 22-AUG-2000: 2000US-0226681.
 PR 22-AUG-2000: 2000US-0226688.
 PR 23-AUG-2000: 2000US-0227182.
 PR 23-AUG-2000: 2000US-0227009.
 PR 30-AUG-2000: 2000US-0228924.
 PR 01-SEP-2000: 2000US-0229287.
 PR 01-SEP-2000: 2000US-0229343.
 PR 01-SEP-2000: 2000US-0229344.
 PR 01-SEP-2000: 2000US-0229345.
 PR 05-SEP-2000: 2000US-0229509.
 PR 05-SEP-2000: 2000US-0229513.
 PR 06-SEP-2000: 2000US-0230437.
 PR 06-SEP-2000: 2000US-0230438.
 PR 08-SEP-2000: 2000US-0231242.
 PR 08-SEP-2000: 2000US-0231243.
 PR 08-SEP-2000: 2000US-0231244.
 PR 08-SEP-2000: 2000US-0231413.
 PR 08-SEP-2000: 2000US-0231414.
 PR 08-SEP-2000: 2000US-0232080.
 PR 08-SEP-2000: 2000US-0232081.
 PR 12-SEP-2000: 2000US-0231968.
 PR 14-SEP-2000: 2000US-0232397.
 PR 14-SEP-2000: 2000US-0232398.
 PR 14-SEP-2000: 2000US-0232399.
 PR 14-SEP-2000: 2000US-0232400.
 PR 14-SEP-2000: 2000US-0232401.
 PR 14-SEP-2000: 2000US-0233063.
 PR 14-SEP-2000: 2000US-0233064.
 PR 14-SEP-2000: 2000US-0233065.
 PR 21-SEP-2000: 2000US-0234223.
 PR 21-SEP-2000: 2000US-0234274.
 PR 25-SEP-2000: 2000US-0234597.
 PR 25-SEP-2000: 2000US-0234598.
 PR 25-SEP-2000: 2000US-0235484.
 PR 27-SEP-2000: 2000US-0235834.
 PR 27-SEP-2000: 2000US-0235836.
 PR 29-SEP-2000: 2000US-0236327.
 PR 29-SEP-2000: 2000US-0236367.
 PR 29-SEP-2000: 2000US-0236368.
 PR 29-SEP-2000: 2000US-0236369.
 PR 29-SEP-2000: 2000US-0236370.
 PR 02-OCT-2000: 2000US-0236802.
 PR 02-OCT-2000: 2000US-0237037.
 PR 02-OCT-2000: 2000US-0237038.
 PR 02-OCT-2000: 2000US-0237039.
 PR 02-OCT-2000: 2000US-0237040.
 PR 13-OCT-2000: 2000US-0239393.
 PR 13-OCT-2000: 2000US-0239397.
 PR 20-OCT-2000: 2000US-0240960.
 PR 20-OCT-2000: 2000US-0241221.
 PR 20-OCT-2000: 2000US-0241785.
 PR 20-OCT-2000: 2000US-0241786.
 PR 20-OCT-2000: 2000US-0241787.

PR 20-OCT-2000: 2000US-0241808.
 PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.
 PR 08-NOV-2000: 2000US-0246528.
 PR 08-NOV-2000: 2000US-0246532.
 PR 08-NOV-2000: 2000US-0246609.
 PR 08-NOV-2000: 2000US-0246610.
 PR 08-NOV-2000: 2000US-0246611.
 PR 08-NOV-2000: 2000US-0246613.
 PR 17-NOV-2000: 2000US-0249207.
 PR 17-NOV-2000: 2000US-0249208.
 PR 17-NOV-2000: 2000US-0249209.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249212.
 PR 17-NOV-2000: 2000US-0249213.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249215.
 PR 17-NOV-2000: 2000US-0249216.
 PR 17-NOV-2000: 2000US-0249217.
 PR 17-NOV-2000: 2000US-0249218.
 PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249264.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251899.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-451936/48.

DR N-PSDB: AAS29661.

Isolated polypeptide for treating, preventing and/or prognosing disorders of the endocrine system such as reproductive disorders, endocrine cancers and also for testing and detection e.g. diagnosis -

Claim 11: SEQ ID No 387; 604pp; English.

Sequences AAU18282-AAU18507 represent endocrine polypeptides of the invention. Endocrine polypeptides and their associated polynucleotides are useful in the diagnosis, treatment and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by determining the presence or absence of a mutation in an endocrine polynucleotide. The treatable disorders include autoimmune diseases such

CC as rheumatoid arthritis, hyperproliferative disorders such as neoplasms
 CC of the breast or liver, cardiovascular disorders such as cardiac arrest,
 CC cerebrovascular disorders such as cerebral ischemia, nervous system
 CC disorders such as Alzheimer's disease, infections caused by bacteria,
 CC viruses and fungi, ocular disorders such as corneal infection, endocrine
 CC disorders such as premature labour and infertility, gastrointestinal
 CC disorders such as Crohn's disease, renal disorders such as
 CC glomerulonephritis and respiratory disorders such as asthma. The
 CC polypeptides can also be used to aid wound healing, to prevent skin aging
 CC due to sunburn, to maintain organs before transplantation, to regenerate
 CC tissues and in chemotaxis. The polypeptides can also be used as a food
 CC additive or preservative to increase or decrease storage capabilities.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Query Match 24.6%; Score 311; DB 22; Length 337;
 Best Local Similarity 33.7%; Pred. No. 8,9e-25;
 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

OY 13 LNKVAATGAASGIGLCAKTLTLDGAKVYLIDREGDKHKIVAE-----GENA 63
 DB 62 LRRAVVVITGATSGLKECKAFYVAGAKLVLCGRNGALEELIRELTASATKVOYTHKP 121
 OY 64 YALQDLFNNQVDNMLADIIELAGLIDIFHANAGAVTGGPVDVDRVLNININA 123
 DB 122 YLVTFPLIDSGAIVAAAEELGCGFYVDILVNNAGISYRTINDTVYDVOKRMENNYG 181
 OY 124 AFRCRAVLPHMIARSGDIIFTSIAGVVPVITWEPYITASKRAVOAFVHTTRQVSQYG 183
 DB 182 PVALTFALLPSMIKROGHVVAISISIOGKMSIFFRSAVYASKHATQAFECLEAEQOYE 241
 OY 184 VRVGAIVLPGPV-----VTALLDMDPK-AKMEALANGSLMQPIEVESVLFVY-TESKNV 236
 DB 242 IEYTVISPGYIHTNLSVNAITADGSKYGVMDTTAOG--RSPVEADVDVLAAGKKKKDV 259
 OY 237 TVRDLV 242
 DB 300 ILADLL 305
 RESULT 10
 AAM42410
 ID AAM42410 standard; Protein; 337 AA.
 AC XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 143.
 XX
 KW Human: nocotropic; antiproliferative; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnecary;
 KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatoprotective; cerebroprotective; antiinflammatory;
 KW antidiabetic; antitubercular; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
 OS Homo sapiens.
 XX
 PN WO200155449-A1.
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01346.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUL-2000; 2000US-0216880.
 PR 14-JUL-2000; 2000US-0218290.
 PR 14-AUG-2000; 2000US-0223447.

PR 01-SEP-2000; 2000US-0229343.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0231243.
 PR 25-SEP-2000; 2000US-0234897.
 PR 29-SEP-2000; 2000US-0236367.
 PR 13-OCT-2000; 2000US-0239937.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246528.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249265.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-476225/51.
 DR N-PSDB; AAI62815.

PT Novel plasma membrane associated proteins useful for diagnosing,
 PT treating, preventing and/or prognosing disorders related to the
 PT proteins, including cancer, immune response and neuronal disorders
 XX
 PS Claim 11; SEQ ID NO 143; 532pp + Sequence Listing; English.

XX The invention relates to novel genes (AAI62752-AAI62961) and proteins
 CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 337 AA;
 Query Match 24.6%; Score 311; DB 22; Length 337;
 Best Local Similarity 33.7%; Pred. No. 8,9e-25;
 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

OY 13 LNKVAATGAASGIGLCAKTLTLDGAKVYLIDREGDKHKIVAE-----GENA 63
 DB 62 LRRAVVVITGATSGLKECKAFYVAGAKLVLCGRNGALEELIRELTASATKVOYTHKP 121
 OY 64 YALQDLFNNQVDNMLADIIELAGLIDIFHANAGAVTGGPVDVDRVLNININA 123
 DB 122 YLVTFPLIDSGAIVAAAEELGCGFYVDILVNNAGISYRTINDTVYDVOKRMENNYG 181
 OY 124 AFRCRAVLPHMIARSGDIIFTSIAGVVPVITWEPYITASKRAVOAFVHTTRQVSQYG 183

Db 182 PVALTALKLPSMIKRRGHIIVAISISIOGKMSIPFRSAVAASKHATQAFDCLRAEMEQYE 241
QY 184 VRGAVLPGPV-----VTALIDDMPK-AKMEALANGSLMQPIEVAESVLFMV-TRSKNV 236
Db 242 IEVTVISPGYIHTNLISYNALITADDSRYGVMDTTTAQG--RSPVEVADVLAAVGKKKKDV 299
QY 237 TVRDLY 242
Db 300 ILADLL 305

RESULT 11
AAB27650
ID AAB27650 standard; Protein; 310 AA.
AC AAB27650;
XX
DT 26-JAN-2001 (first entry)
XX
DE Human protein PRO238.
XX
KW Cardiovascular; endothelial; angiogenic disorder; PRO179;
KW PRO238; PRO364; PRO844; PRO1760; PRO205; PRO321; PRO333;
KW PRO840; PRO877; PRO878; PRO879; PRO882; PRO885; PRO887;
KW gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
XX /label= "Signal peptide"
XX
PN WO200053757-A2.
PD 14-SEP-2000.
XX
PE 24-FEB-2000; 2000MO-US05004.
XX
PR 08-MAR-1999; 99MO-US05028.
PR 12-MAR-1999; 99US-012357.
PR 02-JUN-1999; 99MO-US12252.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99MO-US20111.
PR 15-SEP-1999; 99MO-US21090.
PR 30-NOV-1999; 99MO-US28313.
PR 30-NOV-1999; 99MO-US28409.
PR 02-DEC-1999; 99MO-US28565.
PR 18-FEB-2000; 2000MO-US04342.
PR 22-FEB-2000; 2000MO-US04414.
XX
XX (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Gertsen ME;
PI Goddard A, Gurney AL, Hillan KJ, Marsters SA, Peoni NF, Pitti RM;
PI Wetanabe CK, Williams PM, Wood WJ.
XX
XX WPI; 2000-611444/58.
DR N-PSDB; AAA99902.
XX
PT Novel PRO polypeptides and agonists and antagonists of them, used to
PT diagnose and treat cardiovascular, endothelial and angiogenic disorders
PT
XX
PS Claim 71; Fig 4; 181pp; English.
XX
CC The present invention relates to methods for stimulating or inhibiting
CC angiogenesis and cardiovascularization. The methods involve the use of
CC pharmaceutical compositions based on the following proteins, PRO179,
CC PRO238, PRO364, PRO844, PRO1760, PRO205, PRO321, PRO333,
CC PRO840, PRO877, PRO878, PRO879, PRO882, PRO885 or PRO887. These
CC proteins were identified by isolating cDNA clones encoding secreted
CC proteins. The proteins of the invention may be used to diagnose and

CC treat cardiovascular, endothelial or angiogenic disorders. The present
CC sequence is one of the proteins of the invention.
XX
SQ Sequence 310 AA;
Query Match 24.4%; Score 309; DB 21; Length 310;
Best Local Similarity 33.7%; Pred. No. 1.3e-24;
Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;
QY 13 LNCRAVIAITGASGIGLOCAKTLIDAGAKVYLIDREGDKLHKIVAEI-----GENA 63
Db 35 LRNAVIVITGATISGLGECAKVFYAAGAKVLGCRNGALEBELRELTAHATKRVQTHKP 94
QY 64 YALQDLFPNQVDNMLADILELAGLDIFHANAAYIGGPAVEGDPVDVRLNLINA 123
Db 95 YLVTFEDLTDSGATVAANAELILOCFCGYVDILVNNAGISYRGTIMDTVDVDRKRVETNYFG 154
QY 124 AFRCAVAVLPHMTAQRSGDIIFFTSINGVVPVIMEPIYLTASKFAVQAFVHTTROVQYQ 183
Db 155 PVALTALKLPSMIKRRGHIIVAISISIOGKMSIPFRSAVAASKHATQAFDCLRAEMEQYE 214
QY 184 VRGAVLPGPV-----VTALIDDMPK-AKMEALANGSLMQPIEVAESVLFMV-TRSKNV 236
Db 215 IEVTVISPGYIHTNLISYNALITADDSRYGVMDTTTAQG--RSPVEVADVLAAVGKKKKDV 272
QY 237 TVRDLY 242
Db 273 ILADLL 278

RESULT 12
AAB80238
ID AAB80238 standard; Protein; 310 AA.
XX
AC AAB80238;
XX
DT 24-APR-2001 (first entry)
XX
DE Human PRO238 protein.
XX
KW Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW antiParkinsonian nootropic; neuroprotective; vulnereary; candida;
KW antiangiogenic; vasotropic; antiasthmatic; antineural; cancer;
KW antiarthritic; antiinterfility; antidiabetic; antilyral; diabetes;
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW Ischaemia; inflammation.
XX
OS Homo sapiens.
XX
XX
XX WO200104311-A1.
XX
PD 18-JAN-2001.
XX
PE 22-FEB-2000; 2000MO-US04414.
XX
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 08-SEP-1999; 99MO-US20594.
PR 13-SEP-1999; 99MO-US20944.
PR 15-SEP-1999; 99MO-US21090.
PR 15-SEP-1999; 99MO-US21547.
PR 05-OCT-1999; 99MO-US23089.
PR 29-NOV-1999; 99MO-US28214.
PR 30-NOV-1999; 99MO-US28313.
PR 16-DEC-1999; 99MO-US30095.
PR 20-DEC-1999; 99MO-US30911.
PR 20-DEC-1999; 99MO-US30999.
PR 05-JAN-2000; 99MO-US00219.
XX
XX (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;

PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavich IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumala D,
 PI Williams PM, Wood WI;
 DR WPI: 2001-081051/09.
 DR N-PSDB: AAF72399.
 XX
 PT Sixty one nucleic acids encoding PRO polypeptides which are useful in
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
 PT Alzheimer's disease) -
 PT Alzheimer's disease)
 XX
 PS Claim 1: Fig 56; 393pp; English.
 XX
 CC The present sequence is one of sixty one novel secreted and
 CC transmembrane PRO polypeptides. The PRO polypeptides are
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 CC endometrial bleeding, angiogenesis, ischemias such as coronary
 CC ischemia, atherosclerosis), inflammatory disorders (e.g. asthma,
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
 CC diabetes and retinal disorders such as retinitis pigmentosa.
 CC The PRO nucleic acids have applications in molecular biology, including
 CC use as hybridization probes, and in chromosome and gene mapping.
 CC
 XX
 SQ Sequence 310 AA:
 Query Match 24.4%; Score 309; DB 22; Length 310;
 Best Local Similarity 33.7%; Pred. No. 1.3e-24;
 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;
 QY 13 LNKRAATGASGIGLOCAKTLIDAGAKVYLIDREGDKLHIVAEI-----GENA 63
 DB 35 LKNAVVTGATSGIGLEKAKVYAGAKVILCGRNGALELIRLTASTASHATKQVTHRP 94
 QY 64 VALQDLFENNQGVNMLADIIELAGDIDIFHANAGAYIGPAEGPDVDRVRLNINNA 123
 DB 95 YLVTFDLDSDGAIIVAAAEILLOCFGVLDLVNNAIGSYGTIMDTTVDDKRMENNYTG 154
 QY 124 AFRCYAVLPHMIAORSGDIIFTSSIAGVVPIWEPIYASKFAVOAFVHTTRQVSYQG 183
 DB 155 PVALKKALPSPMIKRRGHIIVAISSIOGKMSIPFRSAVASKHQAQAFDCILAEEMQYE 214
 QY 184 VRVGAVLGPV-----VTRLLDDMPK-AKMEALANGSLMOPIEVAESVLFMW-TRSKAV 236
 DB 215 IEVTVISPEYIHTNLVNAITADSGSRVGYMDTTAOG--RSPVEVAQDVLAIVGKKKKDV 272
 QY 237 TVRDLY 242
 DB 273 ILADLL 278
 XX
 RESULT 13
 ABB95441
 ID ABB95441 standard; Protein: 310 AA.
 XX
 AC ABB95441;
 XX
 DT 19-JUL-2002 (first entry)
 DE Human angiogenesis related protein PRO238 SEQ ID NO: 38.
 XX
 KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
 KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
 KW cardiant; cytoskeletal; antiangiogenic; hypotensive; vulnerable;
 KW antiarteriosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN WO200208284-A2.

XX 31-JAN-2002.
 PD
 XX
 PF 09-JUL-2001; 2001WO-US21735.
 XX
 PR 20-JUL-2000; 2000US-219556P.
 PR 25-JUL-2000; 2000US-220624P.
 PR 25-JUL-2000; 2000US-220664P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 02-AUG-2000; 2000US-222695P.
 PR 17-AUG-2000; 2000US-0643657.
 PR 23-AUG-2000; 2000WO-US23522.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 07-SEP-2000; 2000US-230978P.
 PR 15-SEP-2000; 2000US-000000P.
 PR 18-SEP-2000; 2000US-0664610.
 PR 18-SEP-2000; 2000US-0665350.
 PR 24-OCT-2000; 2000US-242922P.
 PR 08-NOV-2000; 2000US-0709238.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 10-NOV-2000; 2000WO-US30873.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000US-0747259.
 PR 20-DEC-2000; 2000WO-US34956.
 PR 22-JAN-2001; 2001US-0767609.
 PR 28-FEB-2001; 2001US-0796498.
 PR 28-FEB-2001; 2001WO-US06520.
 PR 01-MAR-2001; 2001WO-US06666.
 PR 09-MAR-2001; 2001US-0802706.
 PR 14-MAR-2001; 2001US-0808689.
 PR 22-MAR-2001; 2001US-0816744.
 PR 05-APR-2001; 2001US-0828366.
 PR 10-MAY-2001; 2001US-0854208.
 PR 10-MAY-2001; 2001US-0854280.
 PR 25-MAY-2001; 2001US-0866028.
 PR 25-MAY-2001; 2001US-0866034.
 PR 25-MAY-2001; 2001WO-US17092.
 PR 30-MAY-2001; 2001US-0870574.
 PR 30-MAY-2001; 2001WO-US17443.
 PR 01-JUN-2001; 2001WO-US17692.
 PR 20-JUN-2001; 2001WO-US19800.
 PR 28-JUN-2001; 2001WO-US00000.
 XX
 (GETH) GENENTECH INC.
 PA (BAKE/) BAKER K P.
 PA (FERR/) FERRARA N.
 PA (GERB/) GERBER H.
 PA (GERR/) GERRITSEN M E.
 PA (GODD/) GODDARD A.
 PA (GODO/) GODOWSKI P J.
 PA (GURN/) GURNEY A L.
 PA (HILL/) HILLAN K J.
 PA (MARS/) MARSTERS S A.
 PA (PANU/) PAN J.
 PA (PAON/) PAONI N F.
 PA (STEP/) STEPHAN J F.
 PA (WATA/) WATANABE C K.
 PA (WILL/) WILLIAMS P M.
 PA (WOOD/) WOOD W I.
 XX
 PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 DR WPI: 2002-171999/22.
 DR N-PSDB: ABL95579.
 XX
 PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 XX Claim 11; Fig 38; 567pp; English.

QY	184	VRGAVLPGCV-----VTALLDMPK-AKMEALANGSLMOPLEVAESVLPFW-TRSKNV	236
Db	215	IEYIVISPGIIRHNLNLSNALTADGSRGVAWDITTAQG--RSPVEVAQDVLAVGKKKKDV	272
QY	237	TVRDV 242	
Db	273	ILADLL 278	
RESULT 15			
AAV92513	AAV92513	standard; Protein; 325 AA.	
AC	AAV92513:		
DT	10-AUG-2000	(first entry)	
XX	Human OXRE-10.		
XX	OXRE-10; oxidoreductase; insect-type alcohol dehydrogenase;		
KW	antiproliferative; anticancer; hepatotropic; antiviral; antiasthmatic;		
KW	nootropic; neuroprotectant; antiparkinsonian's; antisclerotic;		
XX	antioxytic; antischizophrenic; anti-inflammatory; antiarthritic.		
XX	Homo sapiens.		
XX	OS		
XX	Key	Location/Qualifiers	
FT	Modified-site	6	
FT	Region	/note="potential phosphorylation site"	
FT	Region	18..37	
FT	Region	/label=transmembrane	
FT	Region	53..243	
FT	Modified-site	/label=signature	
FT	Modified-site	/note="for short-chain dehydrogenase"	
FT	Modified-site	106	
FT	Modified-site	/note="potential phosphorylation site"	
FT	Modified-site	146	
FT	Modified-site	/note="potential phosphorylation site"	
FT	Modified-site	150	
FT	Modified-site	/note="potential phosphorylation site"	
FT	Modified-site	134	
FT	Modified-site	/note="potential phosphorylation site"	
FT	Modified-site	167	
FT	Region	/note="potential phosphorylation site"	
FT	Modified-site	194..222	
FT	Modified-site	/label=signature	
FT	Modified-site	243	
FT	Modified-site	/note="for short-chain dehydrogenase"	
FT	Modified-site	/note="potential glycosylation site"	
FT	Modified-site	268	
FT	Modified-site	/note="potential phosphorylation site"	
XX	WO200020604-A2.		
XX	13-APR-2000.		
XX	06-OCT-1999:	99WO-US2434.	
XX	06-OCT-1998:	98US-0172227.	
XX	02-DEC-1998:	98US-0155202.	
XX	10-MAR-1999:	99US-0123911.	
XX	(INCY-)	INCYTE PHARM INC.	
XX	Lal P, Guegler KJ, Gorgone GA, Corley NC, Baughn MR, Tang YT;		
XX	Hillman JL, Bandman O, Azimzal Y, Au-Young J, Yue H, Lu DAM;		
XX	Yang J;		
XX	WPI: 2000-303785/26.		
XX	N-PDSB; AAA09384.		
XX	Purified polypeptide for treating or preventing disorders associated		

[illegible]

Search completed: March 13, 2003, 16:53:08
Job time : 22.4884 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:37:32 ; Search time 6.82947 Seconds
(without alignments)
1077.057 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHNSVSMNPLNGKVAAL.....TRSKNVTRDLVILPGSYDL 250

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	366	28.9	248	US-09-385-028-11	Sequence 11, Appl
2	319.5	25.3	231	US-09-134-001C-4397	Sequence 4397, Ap
3	309	24.4	350	US-09-149-476-452	Sequence 452, App
4	261	20.6	249	US-09-134-001C-4825	Sequence 4825, Ap
5	255	20.2	246	US-09-238-481-2	Sequence 2, Appl
6	255	20.2	258	US-09-572-810A-2	Sequence 2, Appl
7	247	19.5	258	US-09-504-358-12	Sequence 12, Appl
8	247	19.5	258	US-09-954-314-12	Sequence 12, Appl
9	247	19.5	263	US-09-134-001C-4512	Sequence 12, Appl
10	246.5	19.5	263	US-09-287-097-2	Sequence 12, Ap
11	245.5	19.4	333	US-08-440-856A-4	Sequence 2, Appl
12	242.5	19.2	274	US-09-134-001C-4431	Sequence 4431, Ap
13	242	19.1	262	US-09-363-189B-6	Sequence 6, Appl
14	231.5	18.3	315	US-09-134-001C-5042	Sequence 5042, Ap
15	231.5	18.3	315	US-08-793-035-9	Sequence 9, Appl
16	231.5	18.3	315	US-08-793-035-10	Sequence 9, Appl
17	226.5	17.9	244	US-08-375-962B-13	Sequence 13, Appl
18	226.5	17.9	244	US-08-562-114B-13	Sequence 13, Appl
19	226.5	17.9	244	US-08-729-594A-13	Sequence 13, Appl
20	226.5	17.9	244	US-08-937-993-13	Sequence 13, Appl
21	224	17.7	261	US-09-468-738A-29	Sequence 29, Appl
22	224	17.7	261	US-09-940-019-29	Sequence 29, Appl
23	223.5	17.6	337	US-08-440-856A-3	Sequence 3, Appl
24	222.5	17.5	257	US-09-134-001C-3562	Sequence 3562, Ap
25	221.5	17.1	252	US-08-822-322-8	Sequence 8, Appl
26	216	17.1	252	US-09-466-109-8	Sequence 8, Appl
27	216	17.1	252	US-09-466-109-8	Sequence 8, Appl

28	214	16.9	256	US-08-594-808B-7	Sequence 7, Appl
29	209.5	16.6	263	US-09-134-001C-3505	Sequence 3505, Ap
30	208.5	16.5	243	US-09-239-052-2	Sequence 2, Appl
31	203	16.0	271	US-07-637-865-2	Sequence 2, Appl
32	202	16.0	256	US-09-504-358-14	Sequence 14, Appl
33	202	16.0	256	US-09-954-314-14	Sequence 14, Appl
34	198.5	15.7	313	US-09-413-814-9	Sequence 9, Appl
35	197.5	15.6	292	US-09-468-738A-2	Sequence 2, Appl
36	197.5	15.6	292	US-09-940-019-2	Sequence 2, Appl
37	197.5	15.6	296	US-09-468-738A-23	Sequence 23, Appl
38	197.5	15.6	296	US-09-940-019-23	Sequence 23, Appl
39	194.5	15.4	255	US-08-815-225-4	Sequence 4, Appl
40	190	15.0	186	US-08-858-207A-270	Sequence 270, App
41	189.5	15.0	283	US-09-134-001C-5346	Sequence 5346, Ap
42	189	14.9	301	US-09-288-143-98	Sequence 98, Appl
43	189	14.9	327	US-08-375-962B-12	Sequence 12, Appl
44	189	14.9	327	US-08-562-114B-12	Sequence 12, Appl
45	189	14.9	327	US-08-729-594A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-385-028-11
Sequence 11, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwame A. Aldoo
APPLICANT: Ashish S. Pardekar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jenner Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER TYPE: Floppy disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P57452052
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-385-028-11
Query Match 28.9%; Score 366; DB 4; Length 248;
Best Local Similarity 37.3%; Pred. No. 3.3e-33;

Matches 91: Conservative 41: Mismatches 104: Indels 8: Gaps 3:

QY 9 MNTPLNGKVAITGAASGIGLOCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 65
1 MPSALQGVKVALITGASSIGATFARALAECAVAIAARVEKIRALGDELTAAGAKVHV 60

QY 66 LQIDLFFNNQOVNMLADIIIELAGLDIFPHNAGAVTGPVAGSGPPDVDRNLINNAF 125
61 LELDVADROGVDAVAATVEALGGLDLVNNAGIMLLGPEVADDTTWTMRIDNLLGLM 120

QY 126 RCYRAVLPNMIAGRSQDIITSSAGVVPVIMEDIYTSKFAVOAFVHTTRQVSOXGVR 185
121 YMRALPPLHRSK-GTVVQMSIAGRVNVRNNAVQATKFGVNAFSETLRQETRGVGR 179

QY 186 VGAVLPPEPVYAL---LDMPKAKMEBALANGSLMQPIEAVESVLEFMTRSKNVYVDL 241
180 VVVEPCTTDTTELGHITHTATKEMEYORISQIRKLOADIAEAVRAVYAPHHATYHEI 239

QY 242 YILP 245
240 FIRP 243

RESULT 2
US-09-134-001C-4397
Sequence 4397 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4397
LENGTH: 231
TYPE: PRN
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4397

Query Match 25.3%; Score 319.5; DB 4; Length 231;
Best Local Similarity 30.2%; Pred. No. 5.3e-28;
Matches 70; Conservative 59; Mismatches 92; Indels 11; Gaps 2;

QY 16 KVAITGAASGIGLOCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 75
8 KVAAYTGAASGIGLOCAKTLIDAGAKVLLIDREGDKLHKIVAEI---GENAYA 67

QY 76 VDNMLADIIELAGLDIFPHNAGAVTGPVAGSGPPDVDRNLINNAFRCYRAVLPNMI 135
68 IDMLKAVIDHFGHIDLVNASAGSLSKTDYVNEQMDIMDIVNIGTLHVLAOTPYL 127

QY 136 IAGRSQDIITSSAGVVPVIMEDIYTSKFAVOAFVHTTRQVSOXGVRVAVLP 195
128 LKQSSGHIIINLASVGEPTKTNNAVYCATKAHIAITQSLKELARFGVKVTSISPGMVD 187

QY 196 TALLD--DMKAKMEBALANGSLMQPIEAVESVLEFMTRSKNVYVDL 245
188 TPMTGETDGFERRKLEAQ-----NIDAVVAYLALTOPSHVNVNEVTIRP 230

FILE REFERENCE: P2002PL
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11

1	EARLIER APPLICATION NUMBER: 60/043,671
2	EARLIER FILING DATE: 1997-04-11
3	EARLIER APPLICATION NUMBER: 60/043,674
4	EARLIER FILING DATE: 1997-04-11
5	EARLIER APPLICATION NUMBER: 60/043,669
6	EARLIER FILING DATE: 1997-04-11
7	EARLIER APPLICATION NUMBER: 60/043,312
8	EARLIER FILING DATE: 1997-04-11
9	EARLIER APPLICATION NUMBER: 60/043,313
10	EARLIER FILING DATE: 1997-04-11
11	EARLIER APPLICATION NUMBER: 60/043,672
12	EARLIER FILING DATE: 1997-04-11
13	EARLIER APPLICATION NUMBER: 60/043,315
14	EARLIER FILING DATE: 1997-04-11
15	EARLIER APPLICATION NUMBER: 60/048,974
16	EARLIER FILING DATE: 1997-06-06
17	EARLIER APPLICATION NUMBER: 60/056,886
18	EARLIER FILING DATE: 1997-08-22
19	EARLIER APPLICATION NUMBER: 60/056,877
20	EARLIER FILING DATE: 1997-08-22
21	EARLIER APPLICATION NUMBER: 60/056,889
22	EARLIER FILING DATE: 1997-08-22
23	EARLIER APPLICATION NUMBER: 60/056,893
24	EARLIER FILING DATE: 1997-08-22
25	EARLIER APPLICATION NUMBER: 60/056,630
26	EARLIER FILING DATE: 1997-08-22
27	EARLIER APPLICATION NUMBER: 60/056,878
28	EARLIER FILING DATE: 1997-08-22
29	EARLIER APPLICATION NUMBER: 60/056,662
30	EARLIER FILING DATE: 1997-08-22
31	EARLIER APPLICATION NUMBER: 60/056,872
32	EARLIER FILING DATE: 1997-08-22
33	EARLIER APPLICATION NUMBER: 60/056,882
34	EARLIER FILING DATE: 1997-08-22
35	EARLIER APPLICATION NUMBER: 60/056,637
36	EARLIER FILING DATE: 1997-08-22
37	EARLIER APPLICATION NUMBER: 60/056,903
38	EARLIER FILING DATE: 1997-08-22
39	EARLIER APPLICATION NUMBER: 60/056,888
40	EARLIER FILING DATE: 1997-08-22
41	EARLIER APPLICATION NUMBER: 60/056,879
42	EARLIER FILING DATE: 1997-08-22
43	EARLIER APPLICATION NUMBER: 60/056,880
44	EARLIER FILING DATE: 1997-08-22
45	EARLIER APPLICATION NUMBER: 60/056,894
46	EARLIER FILING DATE: 1997-08-22
47	EARLIER APPLICATION NUMBER: 60/056,911
48	EARLIER FILING DATE: 1997-08-22
49	EARLIER APPLICATION NUMBER: 60/056,636
50	EARLIER FILING DATE: 1997-08-22
51	EARLIER APPLICATION NUMBER: 60/056,874
52	EARLIER FILING DATE: 1997-08-22
53	EARLIER APPLICATION NUMBER: 60/056,910
54	EARLIER FILING DATE: 1997-08-22
55	EARLIER APPLICATION NUMBER: 60/056,864
56	EARLIER FILING DATE: 1997-08-22
57	EARLIER APPLICATION NUMBER: 60/056,631
58	EARLIER FILING DATE: 1997-08-22
59	EARLIER APPLICATION NUMBER: 60/056,845
60	EARLIER FILING DATE: 1997-08-22
61	EARLIER APPLICATION NUMBER: 60/056,892
62	EARLIER FILING DATE: 1997-08-22
63	EARLIER APPLICATION NUMBER: 60/057,761
64	EARLIER FILING DATE: 1997-08-22
65	EARLIER APPLICATION NUMBER: 60/047,595
66	EARLIER FILING DATE: 1997-05-23
67	EARLIER APPLICATION NUMBER: 60/047,599
68	EARLIER FILING DATE: 1997-05-23
69	EARLIER APPLICATION NUMBER: 60/047,588
70	EARLIER FILING DATE: 1997-05-23
71	EARLIER APPLICATION NUMBER: 60/047,585
72	EARLIER FILING DATE: 1997-05-23
73	EARLIER APPLICATION NUMBER: 60/047,586
74	EARLIER FILING DATE: 1997-05-23

EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,590
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,594
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,589
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,593
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,614
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,578
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/043,576
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/047,501
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,670
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/056,632
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,664
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,876
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,881
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,909
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,875
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,862
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,887
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,908
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/048,964
EARLIER	FILING DATE:	1997-06-06
EARLIER	APPLICATION NUMBER:	60/057,650
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/056,884
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/057,669
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/049,610
EARLIER	FILING DATE:	1997-06-13
EARLIER	APPLICATION NUMBER:	60/061,060
EARLIER	FILING DATE:	1997-10-02

Query Match	24.48;	Score 309;	DB 4;	Length 350;
Best Local Similarity	33.78;	Pred. No. 1.5e-26;		
Matches	83;	Conservative 38;	Mismatches 107;	Indels 18; Gaps 5

[illegible]

US-09-504-358-12

Query Match 19.5%; Score 247; DB 4; Length 258;

Best Local Similarity 32.1%; Pred. No. 9.5e-20;

Matches 77; Conservative 35; Mismatches 102; Indels 26; Gaps 6;

QY 13 LNCRAAATGAAGGIGLOCAKTLIDAGAKVYLIDREGDKLHKIYAEIGENAYA---LQL 68

DB 8 LGGDVLVTGAGGIGKATTTALAEGRVVLTDVEDASQVADEVRRMTNGEIRPEPL 67

QY 69 DLFNNQOVNMLADIIELAGGLDI-----FHANAGAYIGGPAESDPVMDRVNLNT 121

DB 68 DVTNP-----AAVTECAKRLDEGMPVYGLMANAGIAPSSSAVDSDLEMLRTVDINL 120

QY 122 NAAPRCVAVLPHMIAQRSDIIFTSIAGVPIWE--PIYASKRVAQAFVHTTRROY 179

DB 121 NGVFWCCREGRKRIANGRSVVTSSIAGFRVSPERHAAYGATKAVALHGLGCEW 180

QY 180 SOYGVRAVLPG---PVVATLIDMPKAKME--EALANGSLMQPIEVAESVLFVWTRS 233

DB 181 AKTGVRNAAVAPGTYTRPIELALKAESPEITSEWTERIPNGRLNDPSEIADGVFLMSNA 240

RESULT 8

US-09-954-314-12

Sequence 12, Application US/09954314

Patent No. 6465224

GENERAL INFORMATION:

APPLICANT: Rouviere, Pierre E.

APPLICANT: Brzostowicz, Patricia C.

TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES

FILE REFERENCE: BC1001 US NA

CURRENT APPLICATION NUMBER: US/09/954,314

PRIOR FILING DATE: 2001-09-17

PRIOR APPLICATION NUMBER: 60/120,702

NUMBER OF SEQ ID NOS: 49

SOFTWARE: Microsoft Office 97

SEQ ID NO 12

LENGTH: 258

TYPE: PR

ORGANISM: Brevibacterium sp HCU

US-09-954-314-12

Query Match 19.5%; Score 247; DB 4; Length 258;

Best Local Similarity 32.1%; Pred. No. 9.5e-20;

Matches 77; Conservative 35; Mismatches 102; Indels 26; Gaps 6;

QY 13 LNCRAAATGAAGGIGLOCAKTLIDAGAKVYLIDREGDKLHKIYAEIGENAYA---LQL 68

DB 8 LGGDVLVTGAGGIGKATTTALAEGRVVLTDVEDASQVADEVRRMTNGEIRPEPL 67

QY 69 DLFNNQOVNMLADIIELAGGLDI-----FHANAGAYIGGPAESDPVMDRVNLNT 121

DB 68 DVTNP-----AAVTECAKRLDEGMPVYGLMANAGIAPSSSAVDSDLEMLRTVDINL 120

QY 122 NAAPRCVAVLPHMIAQRSDIIFTSIAGVPIWE--PIYASKRVAQAFVHTTRROY 179

DB 121 NGVFWCCREGRKRIANGRSVVTSSIAGFRVSPERHAAYGATKAVALHGLGCEW 180

QY 180 SOYGVRAVLPG---PVVATLIDMPKAKME--EALANGSLMQPIEVAESVLFVWTRS 233

DB 181 AKTGVRNAAVAPGTYTRPIELALKAESPEITSEWTERIPNGRLNDPSEIADGVFLMSNA 240

RESULT 9

US-09-134-001C-4512

Sequence 4512, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4512

LENGTH: 263

TYPE: PR

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4512

Query Match 19.5%; Score 247; DB 4; Length 263;

Best Local Similarity 26.7%; Pred. No. 9.8e-20;

Matches 68; Conservative 50; Mismatches 111; Indels 26; Gaps 4;

QY 16 KVAATGAAGGIGLOCAKTLIDAGAKVYLID-----REGDKLHKIYAEIGENAYALQL 68

DB 10 KVAATGAAGGIGLOCAKTLIDAGAKVYLID-----REGDKLHKIYAEIGENAYALQL 65

QY 69 DLFNNQOVNMLADIIELAGGLDI-----FHANAGAYIGGPAESDPVMDRVNLNTNAFRVY 128

DB 66 DVSNRDQVFSVNLQVHEFDLNLVNNAGIGPMTPIESVTPQFQNVGVVNGVFWGI 125

QY 129 RAVLPHM-IAQRSDIIFTSIAGVPIWEPIYASKRVAQAFVHTTRROYSGVRNG 187

DB 126 QALAEOPDKIGHGKIIINATSOAGVGNAGLSTYSKFAVFGITVOAADLAEKNTVN 185

QY 188 AVLPQVATLID-----DPRKAMEALANGSLMQPIEVAESVLFVWTRS 233

DB 186 AFAPGIVETPMKGIKAEKLAENNOQPMWKKOFTDQIALKRLSPEDVANNVSFLAGSD 245

QY 234 KNTVRLVLPQSV 248

DB 246 SDVITGQTIIVDGM 260

RESULT 10

US-09-287-097-2

Sequence 2, Application US/09287097

Patent No. 6255093

GENERAL INFORMATION:

APPLICANT: SCHMUCK, Rainer

APPLICANT: MULLER, Rainer

APPLICANT: WEISSER, Harald

APPLICANT: ENGEL, Alfred

TITLE OF INVENTION: RECOMBINANT MICROBIAL 3-HYDROXYBUTYRATE DEHYDROGENASE,

TITLE OF INVENTION: A PROCESS FOR ITS PRODUCTION AND THE USE THEREOF

FILE REFERENCE: 1614-9007

CURRENT APPLICATION NUMBER: US/09/287,097

CURRENT FILING DATE: 1999-04-07

EARLIER APPLICATION NUMBER: DE/19815685.5

PRIOR FILING DATE: 1997-04-08

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 257

TYPE: PR

ORGANISM: Rhodobacter sphaeroides

US-09-287-097-2

Query Match 19.5%; Score 246.5; DB 4; Length 257;

Best Local Similarity 34.2%; Pred. No. 1.1e-19;

Matches 66; Conservative 33; Mismatches 85; Indels 9; Gaps 3;

QY 13 LNCRAAATGAAGGIGLOCAKTLIDAGAKVYL---IDREGDK--LHKIYAEIGENAYALQ 67

DB 3 LNCRAAATGAAGGIGLOCAKTLIDAGAKVYL---IDREGDK--LHKIYAEIGENAYALQ 62

QY 68 LDFNNQOVNMLADIIELAGGLDI-----FHANAGAYIGGPAESDPVMDRVNLNTNAFRVY 127

Db 63 ADMSDGE---ACRALLETAGCDILVNMAGIOHVSSIEEPPVCKWMAILINLSAFHT 118
Qy 128 VRAVLPHMIAORSGLIETFTSSIAGVAVPIWEPIYASKFAVAQAFVHTTRROVSQYGVAVG 187
Db 119 TAAALPGMRKAKGRIVINIASAGLITSPYKSAVAAKHGVGTTKYTALETAGKGLTCN 178
Qy 188 AVLPGPVVTALLD 200
Db 179 AICPGVLTPLVE 191

RESULT 11
US-08-440-856A-4
Sequence 4, Application US/08440856A
Patent No. 5750873
GENERAL INFORMATION:
APPLICANT: DELLAPORTA, STEPHEN L. METHODS FOR PRODUCING
TITLE OF INVENTION: MATERIALS AND PLANTS WITH SINGLE-SEX FLOWERS
TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE. N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,856A
FILING DATE: 15-MAY-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 05463-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1517
TELEFAX: (202) 887-0763
TELEX: 706141
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-440-856A-4

Query Match 19.4%: Score 245.5; DB 1; Length 333;
Best Local Similarity 27.5%: Pred. No. 2,1e-19;
Matches 77; Conservative 46; Mismatches 117; Indels 40; Gaps 6;
Qy 3 NHSVPSMNTPLNGVAAITGAASIGLQCAKATLLDAGAKVLLIDREGDKLHKIYAEIGEN 62
Db 42 NGAPTTPMKRLEKGVAVITGARGIGEAIVLFLVKGKAKVVIADIDDAAGALAAALGPH 101
Qy 63 AYALQDLFNNQVDNMLADIIELAGGIDIFHANAGAV-----IGPVAEDDPVMDRVL 117
Db 102 VGFPRCVSVVEEDVERAVERAVARYGRDLVCNNAGVLGROTRAASKTSLSPDAGEPDRVL 161
Qy 118 NLN-INAAFCRVAVLPHMIAORSGLIETFTSSIAGVAVPIWEPIYASKFAVAQAFVHTTR 176
Db 162 RVNALGAAALGKHAALA-MTORRAGSIISVAVAGVLCGLGPHAYTASKHAIVGLTKNAA 220
Qy 177 ROVSQYGVAVVLPGPVVTALL-----DDMPKAKMEE 209
Db 221 CELGAHGIRVNCISPFVAVAPMLINAWRGCHADSTADADADADIDLAVPSDQVEKNEE 280

Qy 210 -----ALANGSLMOPIEVAESVLEPMVT-RSKNVTVRDLVI 243
Db 281 VVRGLATLKAGATLRPRDIAEALFLASDDSRYSIGHNLVV 320

RESULT 12
US-09-134-001C-4431
Sequence 4431, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4431
LENGTH: 274
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4431

Query Match 19.2%: Score 242.5; DB 4; Length 274;
Best Local Similarity 28.3%: Pred. No. 3.4e-19;
Matches 71; Conservative 50; Mismatches 113; Indels 17; Gaps 5;

Qy 13 LNKGVAAITGAASIGLQCAKATLLDAGAKVLLIDREGDKLHKIYAEIGEN--AYALQDL 69
Db 6 LENKIAVITGASTGIGASAVALLIEGAHVLLAD-ISDLEETVQSIINDNGKATAVRVD 64
Qy 70 LFNNOQVDNMLADIIELAGGIDIFHANAGAVIG-CGPAEGDPVMDRVNMLINIAARCV 128
Db 65 ISDDKQVKQSEKTAQERGHADVFNNAAGVNDGAGRITHEPVEFDKIMAVDMGTEFLVT 124
Qy 129 RAVLPHMIAORSGLIETFTSSIAGVAVPIWEPIYASKFAVAQAFVHTTRROVSQYGVAVGA 188
Db 125 KFLPLMKMKQ-GGSIINTASFGAADLYRSGVAAAGVINTFKSIAIEYGRINIRANA 183
Qy 189 VLPGPVVTALLDMPKAKMEEA-----LANGSLMOPIEVAESVLEPMVTRSKNVT 237
Db 184 IAPGTIETPLVDNLACTSDEAGQTFRENQKWTPLGRLGTPDEVGLVAFGLASDDSSFI 243
Qy 238 VRDLVILPGSV 248
Db 244 TGETIRIDGV 254

RESULT 13
US-09-363-189B-6
Sequence 6, Application US/09363189B
Patent No. 6242228
GENERAL INFORMATION:
APPLICANT: SUGIYAMA, MASAKAZU
APPLICANT: TONOUCHI, NAOTO
APPLICANT: SUZUKI, SHUNICHI
APPLICANT: YOKOZEKI, KENZO
TITLE OF INVENTION: XLITROL DEHYDROGENASE OF ACETIC ACID BACTERIA AND GENE THERE
FILE REFERENCE: 0010-1024-0
CURRENT APPLICATION NUMBER: US/09/363,189B
CURRENT FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: JP10-216047
PRIOR FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 262
TYPE: PRT
ORGANISM: Glucobacter oxydans

US-09-363-189B-6

Query Match 19.1%; Score 242; DB 4; Length 262;
Best Local Similarity 34.4%; Pred. No. 3.6e-19;
Matches 65; Conservative 28; Mismatches 90; Indels 6; Gaps 3;

QY 9 MNTPLNGVVAITGAASGIGLOCAKTLTDAGAKVVLIDREGDKLHKI--VAELGENAYA 65
DB 1 MSKRFNGKVCVLTGAGNGINGLATLRLAEGTALALDMMNRALREAEKAEVREKYEARS 60
QY 66 LQDLFNNQOVNMLADIIELAGGLDIFHANAGAYIG--GPAEEDPDVMDVNLININA 123
DB 61 YVCDVTSEAVIGTVDSVVRDCKIDLEFNNAG--YOGAPAPQDVPSPDFAVLTITNTG 119
QY 124 AFRCAVAVLPFHIAORSGDIIFTSIAGVVPVIMPEIYASKFAVOAFVHTTRROVSOG 183
DB 120 AFVYKAASROMITONIGYRIVATWASAGVKKPPNMAAGASGALITALTETALDLAPY 179
QY 184 VRVGAVALPQ 192
DB 180 IRVNAISPG 188

RESULT 14
US-09-134-001C-5042

; Sequence 5042, Application US/09134001C
; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5042

; LENGTH: 267

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5042

Query Match 19.1%; Score 241; DB 4; Length 267;
Best Local Similarity 30.6%; Pred. No. 4.8e-19;

Matches 72; Conservative 39; Mismatches 100; Indels 24; Gaps 7;

QY 13 LNKVAATITGAASGIGLOCAKTLTDAGAKVVL---IDREGDKLHKI--VAELGENAYA 66
DB 9 LENKVVLTITGAATIGTIGKSTAEENFGAKAKAVVINYNSDRHSELEKQVYAFGGCTLAV 68
QY 67 QDLFNNQOVNMLADIIELAGGLDIFHANAGAYIGGPAEEDPDVMDVNLININA 126
DB 69 QDVAISEEDIKRMIETITIHFGFLDIINNAGFENSIPTHEMSIDDMQVIVDINLGAVY 128
QY 127 CVRAVLPFHIAO--RSGDIIFTSIAGVVPVIMPEI--YASKFAVOAFVHTTRROVSOG 183
DB 129 GSEETINFLKRNKGTITINISSVHTIP--WPNVYHVAASKGKLIMETSMMEVAYOG 186
QY 184 VRVGAVALPQVVTALDDWPKAMEALANGSLMOP-----EVAESVLFM 229
DB 187 IRINNISPGAIYV---EHTKEFSPPTTREETIKMIPAREIGNADVANAVALFL 237

RESULT 15
US-08-793-035-9

; Sequence 9, Application US/08793035

; Patent No. 6011201

; GENERAL INFORMATION:

; APPLICANT: Stabas, Antoni R.

; APPLICANT: White, Andrew

; APPLICANT: Chase, Dianne
; APPLICANT: Elborough, Keifan
; APPLICANT: Fentem, Philip A.
; TITLE OF INVENTION: B-ketacyl ACP Reductase Genes From
; NUMBER OF INVENTION: Brassica Napus
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX

; COUNTRY: US

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/793,035

; FILING DATE: 28-JUL-1997

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9414622.2

; FILING DATE: 20-JUL-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB95/01678

; FILING DATE: 17-JUL-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Kammerer, Patricia A.

; REGISTRATION NUMBER: 29,775

; REFERENCE/DOCKET NUMBER: MOBT.132

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 713.787.1400

; TELEFAX: 713.787.1440

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 315 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-793-035-9

Query Match 18.3%; Score 231.5; DB 3; Length 315;
Best Local Similarity 28.9%; Pred. No. 7.2e-18;

Matches 67; Conservative 43; Mismatches 111; Indels 11; Gaps 3;

QY 5 SVPSMNTPLNGVVAITGAASGIGLOCAKTLTDAGAKVVL---IDREGDKLHKIYAE 60
DB 66 AVPKVESP---VVVYVTGSRGIGKALISLCKAGCKVLVYNARSKAEAEVSKQIEAVG 121
QY 61 ENAYALQDLFNNQOVNMLADIIELAGGLDIFHANAGAYIGGPAEEDPDVMDVNLIN 120
DB 122 GOAITFGGVNSKEADENAMKTAIDAMGTIDVYVNNAGITRPTLLIRMKKSOMDEVID 181
QY 121 INAAFCVRAVLPFHIAORSGDIIFTSIAGVVPVIMPEIYASKFAVOAFVHTTRROVS 180
DB 182 LTGVFLCTOATKIMMKKKRGRININIASVYGLGINIGQNTYAAKAGVIGFSKTAAREGA 241
QY 181 QYGVRAVAVLPQV---VTALDDWPKAMEALANGSLMOPLEVAESVLFM 229
DB 242 SRINNNVVCPCFIASDMTAKIGEDMEKKILGTIPLGRYGOPEADVAGLVEFL 293

Search completed: March 13, 2003, 16:42:10
Job time: 7.82947 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:12 ; Search time 7.03642 Seconds
(without alignments)
1637.624 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHSHVSPMNTPLNGKVAAL.....TRSKNVTVLDLPGSYDL 250

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PC105_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	100.0	250	9	US-09-802-208B-3
2	311	24.6	337	9	US-09-860-670-143
3	311	24.6	337	10	US-09-764-853-498
4	309	24.4	310	9	US-09-905-291A-153
5	309	24.4	310	9	US-09-902-853-153
6	309	24.4	310	9	US-09-907-824-153
7	309	24.4	310	9	US-09-907-841-153
8	309	24.4	310	9	US-09-904-011-153
9	309	24.4	310	9	US-09-906-742-153
10	309	24.4	310	9	US-09-906-838-153
11	309	24.4	310	9	US-09-907-613-153
12	309	24.4	310	9	US-09-907-942-153
13	309	24.4	310	9	US-09-904-820-153
14	309	24.4	310	9	US-09-904-859-153
15	309	24.4	310	9	US-09-909-204-153
16	309	24.4	310	9	US-09-904-786-153
17	309	24.4	310	9	US-09-906-646-153
18	309	24.4	310	9	US-09-906-700-153
19	309	24.4	310	9	US-09-902-903-153

20	309	24.4	310	9	US-09-903-749A-153	Sequence 153, App
21	309	24.4	310	9	US-09-903-786-153	Sequence 153, App
22	309	24.4	310	9	US-09-909-330-153	Sequence 153, App
23	309	24.4	310	10	US-09-909-088B-153	Sequence 153, App
24	291.5	23.0	254	9	US-09-978-758-2	Sequence 2, Appl1
25	281.5	22.3	203	10	US-09-823-901-7	Sequence 7, Appl1
26	278	22.0	206	9	US-09-997-816-4	Sequence 4, Appl1
27	270.5	21.4	251	10	US-09-815-242-5689	Sequence 5689, Ap
28	270.5	21.4	272	10	US-09-815-242-12688	Sequence 12688, A
29	255	20.2	246	10	US-09-815-242-5461	Sequence 5461, Ap
30	255	20.2	246	10	US-09-815-242-12123	Sequence 12123, A
31	255	20.2	246	10	US-09-815-242-12803	Sequence 12803, A
32	255	20.2	246	10	US-09-815-242-13100	Sequence 13100, A
33	251	19.8	263	10	US-09-815-242-14092	Sequence 14092, A
34	250.5	19.8	266	9	US-09-981-353-40	Sequence 40, Appl
35	250.5	19.8	266	10	US-09-931-186-18	Sequence 18, Appl
36	250.5	19.8	276	9	US-09-944-160-24	Sequence 24, Appl
37	249.5	19.7	266	9	US-09-885-303A-22	Sequence 22, Appl
38	249	19.7	261	10	US-09-815-242-5813	Sequence 5813, Ap
39	247	19.5	258	10	US-09-954-314-12	Sequence 12, Appl
40	245	19.4	253	10	US-09-815-242-11842	Sequence 11842, A
41	242	19.1	262	10	US-09-802-853-6	Sequence 6, Appl1
42	242	19.1	312	9	US-09-885-303A-16	Sequence 16, Appl
43	241.5	19.1	278	9	US-10-033-245-2	Sequence 2, Appl1
44	241.5	19.1	278	9	US-10-033-223-2	Sequence 2, Appl1
45	241.5	19.1	278	9	US-10-033-167-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-09-802-208B-3
Sequence 3, Application US/09802208B
Publication No. US20030041352A1
GENERAL INFORMATION:
APPLICANT: Parrott, Wayne
APPLICANT: Lafayette, Peter
APPLICANT: Kane, Patrick
TITLE OF INVENTION: Antibiotol or Ribitol As Positive Selectable Markers
FILE REFERENCE: UGA 855R
CURRENT APPLICATION NUMBER: US/09/802,208B
CURRENT FILING DATE: 2001-03-08
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 250
TYPE: PRT
ORGANISM: Escherichia coli
US-09-802-208B-3

Query Match	100.0%	Score 1265;	DB 9;	Length 250;
Best Local Similarity	100.0%	Pred. No. 1.2e-117;		
Matches 250;	Conservative	0;	Mismatches 0;	Indels 0;
			Gaps 0;	
Oy	1	MNHSHVSPMNTPLNGKVAALITGAASIGLOCAKTLTLDAGAKVVLIDREGDKLHKIYAELG	60	
Db	1	MNHSHVSPMNTPLNGKVAALITGAASIGLOCAKTLTLDAGAKVVLIDREGDKLHKIYAELG	60	
Oy	61	ENAYALQDLFPNQOVNMLADIIELAGGIDIFHANAGAYIGGPAVEGDPDWDRLNLN	120	
Db	61	ENAYALQDLFPNQOVNMLADIIELAGGIDIFHANAGAYIGGPAVEGDPDWDRLNLN	120	
Oy	121	INAAFCVRAVLPHMTAORSSGDIIFTSISAGVVPVWEPYATSKRAVQAFVHTTRRVS	180	
Db	121	INAAFCVRAVLPHMTAORSSGDIIFTSISAGVVPVWEPYATSKRAVQAFVHTTRRVS	180	
Oy	181	OYGVRAVGLPGPVYATLDDMPKAKMEBALANGSLMOPIEVAESVLFMTTRSKNTVVD	240	
Db	181	OYGVRAVGLPGPVYATLDDMPKAKMEBALANGSLMOPIEVAESVLFMTTRSKNTVVD	240	
Oy	241	LVILPGSYDL 250		

Db 241 LVLPSPVDL 250

RESULT 2

US-09-860-670-143

; Sequence 143, Application US/09860670

; Patent No. US20020165137A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA127P1

; CURRENT APPLICATION NUMBER: US/09/860,670

; CURRENT FILING DATE: 2001-05-21

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 289

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 143

; LENGTH: 337

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-860-670-143

Query Match

Best Local Similarity 33.7%; Score 311; DB 9; Length 337;

Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

QY 13 LNCRAVAITGASGIGLOCAKTLDDAGAKVVLIDREGDKLHKIVAEI-----GENA 63

Db 62 LRNAVVTIGATISGLGKCAKVFYAAGAKVLCGRNGALEELIRELTASHATKVTGTHKP 121

QY 64 YALQDLFNNQOVNMLADIIELAGGIDIFHANAGAVIGGPAEGDPVDVRLNLTINA 123

Db 122 YLVTFDLIDSGAIVAAAEELIQCFCGYVDILVNNAGISYRGTIMDTVDVKRVMEETNYFG 181

QY 124 AFRCVRAVLEPHMTAORSGLDIFTSIAGVVPVIMEPIYTSKRAVQAFTVTTROVSQYG 183

Db 182 PVALTALPLSPMKRQGHIVAISISIOGKMSIFRSAYASKHATQAFEDCLRAEMQYE 241

QY 184 VRGAVLPGCV-----VTALLDDMPK-AKMEALANGSLMOPIEVAESVLFMV-TRSKNV 236

Db 242 IEVTVISPGYIHNLSVNAITADGSRGVMDTTTAAQ--RSPVEVADVDVLAAGKKKKDV 299

QY 237 TVRDLY 242

Db 300 ILADLL 305

RESULT 3

US-09-764-853-498

; Sequence 498, Application US/09764853

; Patent No. US20020090672A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PJ206

; CURRENT APPLICATION NUMBER: US/09/764,853

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 939

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 498

; LENGTH: 337

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-853-498

Query Match

Best Local Similarity 33.7%; Score 311; DB 10; Length 337;

Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

QY 13 LNCRAVAITGASGIGLOCAKTLDDAGAKVVLIDREGDKLHKIVAEI-----GENA 63

Db 62 LRNAVVTIGATISGLGKCAKVFYAAGAKVLCGRNGALEELIRELTASHATKVTGTHKP 121

QY 64 YALQDLFNNQOVNMLADIIELAGGIDIFHANAGAVIGGPAEGDPVDVRLNLTINA 123

Db 122 YLVTFDLIDSGAIVAAAEELIQCFCGYVDILVNNAGISYRGTIMDTVDVKRVMEETNYFG 181

QY 124 AFRCVRAVLEPHMTAORSGLDIFTSIAGVVPVIMEPIYTSKRAVQAFTVTTROVSQYG 183

Db 182 PVALTALPLSPMKRQGHIVAISISIOGKMSIFRSAYASKHATQAFEDCLRAEMQYE 241

QY 184 VRGAVLPGCV-----VTALLDDMPK-AKMEALANGSLMOPIEVAESVLFMV-TRSKNV 236

Db 242 IEVTVISPGYIHNLSVNAITADGSRGVMDTTTAAQ--RSPVEVADVDVLAAGKKKKDV 299

QY 237 TVRDLY 242

Db 300 ILADLL 305

RESULT 4

US-09-905-291A-153

; Sequence 153, Application US/09905291A

; Patent No. US20020160374A1

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.

; APPLICANT: Ashkenazi, Avi

; APPLICANT: Botstein, David

; APPLICANT: Desnovers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerber, Hanspeter

; APPLICANT: Gertlisen, Mary E.

; APPLICANT: Goddard, A.

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth, J.

; APPLICANT: Kijavlin, Ivar J.

; APPLICANT: Mather, Jennie P.

; APPLICANT: Pan, James

; APPLICANT: Paonli, Nicholas F.

; APPLICANT: Roy, Margaret Ann

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William, I.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE REFERENCE: 10466-14

; CURRENT APPLICATION NUMBER: US/09/905,291A

; CURRENT FILING DATE: 2001-07-12

; PRIOR APPLICATION NUMBER: PCT/US00/04414

; PRIOR FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 60/143,048

; PRIOR FILING DATE: 1999-07-07

; PRIOR APPLICATION NUMBER: US 60/145,698

; PRIOR FILING DATE: 1999-07-26

; PRIOR APPLICATION NUMBER: US 60/146,222

; PRIOR FILING DATE: 1999-07-28

; PRIOR APPLICATION NUMBER: PCT/US99/20594

; PRIOR FILING DATE: 1999-09-08

; PRIOR APPLICATION NUMBER: PCT/US99/20944

; PRIOR FILING DATE: 1999-09-13

; PRIOR APPLICATION NUMBER: PCT/US99/21090

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/21547

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/23089

; PRIOR FILING DATE: 1999-10-05

; PRIOR APPLICATION NUMBER: PCT/US99/28214

; PRIOR FILING DATE: 1999-11-29

```

APPLICANT: Tumms, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/902,853
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US/09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 153
LENGTH: 310
TYPE: PRT
ORGANISM: Homo Sapien
US-09-902-853-153

Query Match      24.4% Score 309; DB 9; Length 310;
Best Local Similarity | 33.7%; Pred. No. 7,4e-23;
Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps

OY 13 LNKGAATITGASGIGLCACAKTLIDAGAKVYLIDEQDKLHKIVAEI-----GENA 63
    |   |||||:::|||||:::|||||:::|   |:::|
Db 35 LRNAVVTITGATSGIGKCKCAKYFAAGAKIVLCGNNGALELEILNETSHATKVQTINP 94
    |   :|:::|:::|:::|:::|:::|:::|   |:::|

OY 64 YALOLDLFNNOOVDMNLADIETFLAGGLDIFHNAGAYTGGPYAEGDPDYMDRYLNININA 123
    |   ::::|:::|:::|:::|:::|:::|   |:::|
Db 95 YLVTFEDLTDSGAIFYAAAEIILOCFCGYVDILVNNAGISYGTIMDTTVDDVKRMETNYFG 154
    |   ::::|:::|:::|:::|:::|:::|   |:::|

OY 124 AFRCRAVLPHMIAQRSDIIFTSSIAGVPYIMEPIYTASKFAFYAOAFVHTTRROVSOG 183
Db 155 PVALTALKPLPKIKRCHGVIAISSIQGRMSIPFSAATAASHAHATOFADDCIRAEEMOTE 214
    |   ::::|:::|:::|:::|:::|:::|   |:::|

OY 184 VRVGAVLEGPV---VVALLDDMKR-AKMEBALANGSLMDPIEVAESVLYFW-TRSKNV 236
    |   ::::|:::|:::|:::|:::|:::|   |:::|
Db 215 IEVYISPGYIHTNLSVNAITADGSRYGVMOTTTAQ--RSPVEVAQDVLAIVGKKKDV 272
OY 237 TVRDLY 242
    |   :::|
Db 273 ILADLL 278

```

RESULT 6
US-09-907-824-153
Sequence 153, Application US/09907824
Publication No. US20020197671A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mathier, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,824
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/22089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423

SEQ ID NO 153
LENGTH: 310
TYPE: PRT
ORGANISM: Homo Sapien
US-09-907-824-153
Query Match 24.48; Score 309; DB 9; Length 310;
Best Local Similarity 33.78; Pred. No. 7,4e-23;
Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;
Y 13 LNKVAAITGAAGSIGGICGACTLLDAGAKVYLIDREGDKLHKIYAEI-----GENA 63
D 35 LRNAVVTTCATSGCECAKVFYAAGKVLGCRNGALBELRELTAASHATVQTHKP 94
Y 64 YALQDLFNNQVDNMLADIETLAGLDIFRANAGAYIGFVAGDDPDVMDRVNLNINA 123
D 95 YLVTFEDLDSGAIYAAAEIQCFCGYVDILVNNAGISYRGIMDTYDVKRWETVFG 154
Y 124 AFRCRAVLPHMIAQRSGDILETSSINAGVVPVIMEPIYTSKFAVQAFVHTTRQVSQYG 183
D 155 PVALTAKLLPSMIKRRGHIYAIISSIOGKMSIPRSAYVAASKHATQAFDFCLRAEMEQYE 214
Y 184 VRGAVLPGPY-----VTALDDMPK-AKMEALANGSLMQPIEVAESVLFMV-TRSKNV 236
D 215 LEVTVISPGYIHTNLVSNAITADGSRGVMDTTAAG--RSPVEADVDLAAGKKKKDV 272
Y 237 TVRDV 242
D 273 ILADLL 278
RESULT 7
US-09-907-841-153
Sequence 153, Application US/09907841
Publication No. US20020198366A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mathier, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,841
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222

1	APPLICANT:	Genentech, Inc.
2	APPLICANT:	Ashtenzai, Avi
3	APPLICANT:	Borstein, David
4	APPLICANT:	Desnoyers, Luc
5	APPLICANT:	Eaton, Dan L.
6	APPLICANT:	Ferrara, Napoleone
7	APPLICANT:	Filvaroff, Ellen
8	APPLICANT:	Fong, Sherman
9	APPLICANT:	Gao, Wei-qiang
10	APPLICANT:	Gerber, Hanspeter
11	APPLICANT:	Gertsen, Mary E.
12	APPLICANT:	Goddard, A.
13	APPLICANT:	Godowski, Paul J.
14	APPLICANT:	Grimaldi, Christopher J.
15	APPLICANT:	Gurney, Austin L.
16	APPLICANT:	Hillan, Kenneth, J.
17	APPLICANT:	Kiljavin, Ivar J.
18	APPLICANT:	Maheer, Jeanne P.
19	APPLICANT:	Pan, James
20	APPLICANT:	Pooni, Nicholas F.
21	APPLICANT:	Roy, Margaret Ann
22	APPLICANT:	Stewart, Timothy A.

QY 13 L N G V A A T T G A A S I G I O C A K T L L D A G K A V V L I R E D G K L H K I A E L ----- GENA 63
 Db 35 L R N A V V I T G A T S G I G K E C A K V F P A A G A K E V L C R G N G A L E B E L I R E L T A S H A T K V O T H N K 94
 QY 64 Y A L O L D L E N N O V M L A D I I E L A G L D I F I A N N A G I I G G V A G D D P W M D R V L N L I N A 122
 Db 95 Y L V F I F D L D S A I A A A A E I I O C F G V D I L N N A G I S T R G I T M T V D V D K R V E T H P F G 154
 QY 124 A F R C V A V L P H M I A O R S D I I F T S S I A G V V F V I E P Y T A S K F A V Q A F V H T T R O V S O Y G 183
 Db 155 P V A L T K A L L P M I R K R O C H I A I V A I S I O G K M I P E R S A V A A S N A T O A F F C D L R A E M O Y E 214
 QY 184 V R V G A V L E G P V ----- V T A L L D M P K - A K M E A L A N S I M O P I V A S V L E M V - T R S K N V 236
 Db 215 I E V V I A S I G Y I H T M L S V A I T A D S R G V M D T T T A A G - R S P V E A D V D L A A G K K K K D V 272
 QY 237 T V R D L V 242
 : : I I :

Db 273 ILADL 278

RESULT 9

```
; Sequence 153, Application US/09906742
; Publication No. US20030023034A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,742
; PRIOR APPLICATION NUMBER: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
```

```
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 153
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-906-742-153
```

Query Match 24.4%; Score 309; DB 9; Length 310;
Best Local Similarity 33.7%; Pred. No. 7,4e-23;
Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;

```
QY 13 LNKGVAAITGAASGIGQCAKTLIDAGAKVYLIDREGDKLKKIYAEI-----GENA 63
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35 LRNAVVTGATSGIKCEKACAFYAAGAKVLICRNGALBELIRELTASHATKVTQTHKP 94
QY 64 YALDLLENQOVNDMLDIETLAGSDIFPANAAGVIYGVAEGDDVDMDRVNLNINA 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 YLTFEDLDSGAIYAAAEIILQCFGYVDILVNNAGISYRGITMDTVVDKRVMEIYFG 154
QY 124 AFRCVRAVLPHMIQRSQDIETSIAGVVPVIEPIYTASKFAVQAFVHTTROVSQYG 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 PVALTAKLPLPSMIKRQGHIVAISISIQKMSIPRSVAASKHATQAFPODLRAEMQYE 214
QY 184 VRGAVLPGPV-----VTALIDMPK--AKMEALANGSLMOPIVASVLFMV--TRSKNV 236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 215 IEVTVISPGYIHTNLVNAITADSGRYGVMDTTAOG--RSPVEVDVLAAGKKKKKV 272
QY 237 TVRDLV 242
      |||||
Db 273 ILADL 278
```

RESULT 10

```
; Sequence 153, Application US/09906638
; Publication No. US20030027143A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,838
; PRIOR APPLICATION NUMBER: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
```

[illegible]

```

1  APPLICANT: Ferrara, Napoleone
2  APPLICANT: Filvaroff, Ellen
3  APPLICANT: Fong, Sherman
4  APPLICANT: Gao, Wei-Qiang
5  APPLICANT: Gerber, Hanspeter
6  APPLICANT: Gerltsen, Mary E.
7  APPLICANT: Goddard, A.
8  APPLICANT: Godowski, Paul J.
9  APPLICANT: Grimaldi, Christopher J.
10 APPLICANT: Gurney, Austin L.
11 APPLICANT: Hillan, Kenneth, J.
12 APPLICANT: Kljavin, Ivar J.
13 APPLICANT: Mather, Jenote P.
14 APPLICANT: Pan, James
15 APPLICANT: Paoni, Nicholas F.
16 APPLICANT: Roy, Margaret Ann
17 APPLICANT: Stewart, Timothy A.
18 APPLICANT: Tumas, Daniel
19 APPLICANT: Williams, P. Mickey
20 APPLICANT: Wood, William, I.
21 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
22 FILE REFERENCE: 10466-14
23 CURRENT APPLICATION NUMBER: US/09/907,613
24 PRIOR FILING DATE: 2001-07-17
25 PRIOR APPLICATION NUMBER: PCT/US00/04414
26 PRIOR FILING DATE: 2000-02-22
27 PRIOR APPLICATION NUMBER: US 60/143,048
28 PRIOR FILING DATE: 1999-07-07
29 PRIOR APPLICATION NUMBER: US 60/145,698
30 PRIOR FILING DATE: 1999-07-26
31 PRIOR APPLICATION NUMBER: US 60/146,222
32 PRIOR FILING DATE: 1999-07-28
33 PRIOR APPLICATION NUMBER: PCT/US99/20594
34 PRIOR FILING DATE: 1999-09-08
35 PRIOR APPLICATION NUMBER: PCT/US99/20944
36 PRIOR FILING DATE: 1999-09-13
37 PRIOR APPLICATION NUMBER: PCT/US99/21090
38 PRIOR FILING DATE: 1999-09-15
39 PRIOR APPLICATION NUMBER: PCT/US99/21547
40 PRIOR FILING DATE: 1999-09-15
41 PRIOR APPLICATION NUMBER: PCT/US99/23089
42 PRIOR FILING DATE: 1999-10-05
43 PRIOR APPLICATION NUMBER: PCT/US99/28214
44 PRIOR FILING DATE: 1999-11-29
45 PRIOR APPLICATION NUMBER: PCT/US99/28313
46 PRIOR FILING DATE: 1999-11-30
47 PRIOR APPLICATION NUMBER: PCT/US99/28564
48 PRIOR FILING DATE: 1999-12-02
49 PRIOR APPLICATION NUMBER: PCT/US99/28565
50 PRIOR FILING DATE: 1999-12-02
51 PRIOR APPLICATION NUMBER: PCT/US99/30095
52 PRIOR FILING DATE: 1999-12-16
53 PRIOR APPLICATION NUMBER: PCT/US99/30911
54 PRIOR FILING DATE: 1999-12-20
55 PRIOR APPLICATION NUMBER: PCT/US99/30999
56 PRIOR FILING DATE: 1999-12-20
57 PRIOR APPLICATION NUMBER: PCT/US00/00219
58 PRIOR FILING DATE: 2000-01-05
59 NUMBER OF SEQ ID NOS: 423
60 SEQ ID NO 153
61 LENGTH: 310
62 TYPE: PRT
63 ORGANISM: Homo sapiens
64 US-09-907-613-153
65
66 Query Match 24.4%; Score 309; DB 9; Length 310;
67 Best Local Similarity 33.7%; Pred. No. 7.4e-23;
68 Matches 83; Conservative 38; Mismatches 107; Indels 18; Gaps 5;
69
70 13 LMGKAATTGCAAGTGICGACAKTLIDNAGKAVYLIDEGDKLHIVAEI-----GENA 63
71      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
72 35 LRNAVVVITGATGATGKCAKVFYAAAGKLVLCGRNGCALLEELIELRLASHTKVTQTHKP 94

```

```

OY      64 YALDIDJFNNOQVNMJLADIIEIETLAGGDIFHANAGAVIGPVAEGOPVDVRLNIN 123
Db      95 YLVFIEDLDSALYIAAAAEIIOCTGYDILYNNNGISYRGTIMDYVDKRMETNYG 154
OY      124 AFRCRAVLPHMIAORSGDIIETTSIAGVVPVIMEPIYTSKRAVOAFVHTTRQVSQY 183
Db      155 PVALTKALPIMIKRQGHIVAIISIGKMSIPRSAYSAKSHATQAFPCLEAMEQYE 214
OY      184 VRGCVLPGPV-----VTALLDMPK AKHEEALANSIMQPIEVAESYFMY-TSKYN 236
Db      215 IEVVISGVIHTNLSVNAITGDSRGVMDTTTAQG--RSPEVADVDYLAAGKRRKDV 272
OY      237 TVRDLV 242
Db      273 ILADLL 278

RESULT 12
US-09-907-942-153
: Sequence 153, Application US/09907942
: Publication No. US20030027146A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertselsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paonl, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,942
CURRENT FILING DATE: 2002-01-22
PRIORITY APPLICATION NUMBER: PCT/US00/04414
PRIORITY FILING DATE: 2000-02-22
PRIORITY APPLICATION NUMBER: US 60/143,048
PRIORITY FILING DATE: 1999-07-07
PRIORITY APPLICATION NUMBER: US 60/145,698
PRIORITY FILING DATE: 1999-07-26
PRIORITY APPLICATION NUMBER: US 60/146,222
PRIORITY FILING DATE: 1999-07-28
PRIORITY APPLICATION NUMBER: PCT/US99/20594
PRIORITY FILING DATE: 1999-09-08
PRIORITY APPLICATION NUMBER: PCT/US99/20944
PRIORITY FILING DATE: 1999-09-13
PRIORITY APPLICATION NUMBER: PCT/US99/21090
PRIORITY FILING DATE: 1999-09-15
PRIORITY APPLICATION NUMBER: PCT/US99/21547
PRIORITY FILING DATE: 1999-09-15
PRIORITY APPLICATION NUMBER: PCT/US99/23089
PRIORITY FILING DATE: 1999-10-05
PRIORITY APPLICATION NUMBER: PCT/US99/28214
PRIORITY FILING DATE: 1999-11-29

```

	PRIOR APPLICATION NUMBER:	PCT/US99/28313	
	PRIOR FILING DATE:	1999-11-30	
	PRIOR APPLICATION NUMBER:	PCT/US99/28554	
	PRIOR FILING DATE:	1999-12-02	
	PRIOR APPLICATION NUMBER:	PCT/US99/28565	
	PRIOR FILING DATE:	1999-12-02	
	PRIOR APPLICATION NUMBER:	PCT/US99/30095	
	PRIOR FILING DATE:	1999-12-16	
	PRIOR APPLICATION NUMBER:	PCT/US99/30911	
	PRIOR FILING DATE:	1999-12-20	
	PRIOR APPLICATION NUMBER:	PCT/US99/30999	
	PRIOR FILING DATE:	1999-12-20	
	PRIOR APPLICATION NUMBER:	PCT/US00/00219	
	PRIOR FILING DATE:	2000-01-05	
	NUMBER OF SEQ ID NOS:	423	
	SEQ ID NO 153		
	LENGTH:	310	
	TYPE:	PRF	
	ORGANISM:	Homo sapiens	
	US-09-907-942-153		
	Query Match	24.4%; Score 309; DB 9; Length 310;	
	Best Local Similarity	33.7%; Pred. No. 7.4e+23;	
	Matches	83; Conservative 38; Mismatches 107; Indels 18; Gaps 5.	
OY	13	LNGKVAATGAASGIGLOCAKTLLDAGAKVVLIDREGDKLHKIVAEI-----GENA 63	
DB	35	LRNAVVTGTGATSGIGKBCAKFYAAAGAKLYLCGGNNGALELEILRELTASHATKYQTHRP 94	
OY	64	VALADLDLFNNQOVDMMLDIIELAGGLDFPHNAGAYTGGPAVEGRDPVMDLVNLNTINA 123	
DB	95	YLVTDTDLDSGAIYAAAEEIILOCFGVDILVNNAISITRGTTIMDTTVDDKKVMETNIFG 154	
OY	124	AFCRCRAVLPHMIAQRSDDIFTSSIAGVPVIWEPIYTASKFAVOAFVHTTRROVSOGX 183	
DB	155	PVALTKALLPSMIKRQGHIVAIISSIQGMKSIPFRSAVAASKHATQAFFDCIRAMEOYE 214	
OY	184	VRVGVAVLGPRV----VALALDDMKP-AKMEBALNGSLMOPIEVAESYLEFW-TRSKV 236	
DB	215	LEVYVISGYIHNTLSVNAITRADSGRYGCVMTTAAQG--RSPFEVAQDVLAAGKKKRDV 272	
OY	237	TVRDLV 242	
DB	273	ILADLL 278	
	RESULT 13		
	US-09-904-820-153		
	: Sequence 153, Application US/09904820		
	: Publication No. US20030036094A1		
	GENERAL INFORMATION:		
	APPLICANT: Genentech, Inc.		
	APPLICANT: Ashkenazi, Avi		
	APPLICANT: Botstein, David		
	APPLICANT: Desnoyers, Luc		
	APPLICANT: Eaton, Dan L.		
	APPLICANT: Ferrara, Napoleon		
	APPLICANT: Filvaroff, Ellen		
	APPLICANT: Fong, Sherman		
	APPLICANT: Gao, Wei-Qiang		
	APPLICANT: Gerber, Hanspeter		
	APPLICANT: Gerritsen, Mary E.		
	APPLICANT: Goddard, A.		
	APPLICANT: Godowski, Paul J.		
	APPLICANT: Grimaldi, Christopher J.		
	APPLICANT: Gurney, Austin L.		
	APPLICANT: Hillan, Kenneth, J.		
	APPLICANT: Kijavini, Ivar J.		
	APPLICANT: Mather, Jennie P.		
	APPLICANT: Pan, James		
	APPLICANT: Paonl, Nicholas F.		
	APPLICANT: Roy, Margaret Ann		
	APPLICANT: Stewart, Timothy A.		


```

; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 153
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-859-153 .

```

Query Match	24.4%	Score 309;	DB 9;	length 310;
Best Local Similarity	33.7%	Pred. No. 7.4e-23;		
Matches	83;	Conservative	38;	Mismatches 107;
				Indels 18;
				Gaps 5.

[illegible]

RESULT 15
US-09-909-204-153
Sequence 153, Application US/0909204
Publication No. US20030036061A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashtenazl, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavan, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US/09/909,204
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698

```

? PRIOR FILING DATE: 1999-07-26
? PRIOR APPLICATION NUMBER: US 60/146,222
? PRIOR FILING DATE: 1999-07-28
? PRIOR APPLICATION NUMBER: PCT/US99/20594
? PRIOR FILING DATE: 1999-09-08
? PRIOR APPLICATION NUMBER: PCT/US99/20944
? PRIOR FILING DATE: 1999-09-13
? PRIOR APPLICATION NUMBER: PCT/US99/21090
? PRIOR FILING DATE: 1999-09-15
? PRIOR APPLICATION NUMBER: PCT/US99/21547
? PRIOR FILING DATE: 1999-09-15
? PRIOR APPLICATION NUMBER: PCT/US99/23089
? PRIOR FILING DATE: 1999-10-05
? PRIOR APPLICATION NUMBER: PCT/US99/28214
? PRIOR FILING DATE: 1999-11-29
? PRIOR APPLICATION NUMBER: PCT/US99/28313
? PRIOR FILING DATE: 1999-11-30
? PRIOR APPLICATION NUMBER: PCT/US99/28564
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US99/28565
? PRIOR FILING DATE: 1999-12-02
? PRIOR APPLICATION NUMBER: PCT/US99/30095
? PRIOR FILING DATE: 1999-12-16
? PRIOR APPLICATION NUMBER: PCT/US99/30911
? PRIOR FILING DATE: 1999-12-20
? PRIOR APPLICATION NUMBER: PCT/US99/30999
? PRIOR FILING DATE: 1999-12-20
? PRIOR APPLICATION NUMBER: PCT/US00/00219
? PRIOR FILING DATE: 2000-01-05
? NUMBER OF SEQ ID NOS: 423
? SEQ ID NO 153
? LENGTH: 310
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-909-204-153

```

Query Match	24.48;	Score 309;	DB 9;	Length 310;
Best Local Similarity	33.78;	Pred. No. 7.4e-23;		
Matches	83;	Conservative	38;	Mismatches 107;
			Indels	18;
			Gaps	5

[illegible]

```
Search completed: March 13, 2003, 16:51:21
Job time : 9.03642 secs
```

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:50:47 ; Search time 8.69205 Seconds

(without alignments)
2765.009 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MNHNSVPSMNTPLNGKVAAL.....TRSKNVTVRDVILPGSVDL 250

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1100	87.0	249	1 S07134	ribitol 2-dehydrog
2	1073	84.8	247	1 DEKBR	ribitol 2-dehydrog
3	843	66.6	242	2 AC3632	ribitol 2-dehydrog
4	573	45.3	243	2 B98199	ribitol 2-dehydrog
5	573	45.3	243	2 AG3087	ribitol 2-dehydrog
6	333	26.3	249	2 E95283	probable [imported
7	331	26.2	231	2 G90050	hypothetical prote
8	304	24.0	325	2 T17307	hypothetical prote
9	295.5	23.4	255	2 T44991	oxidoreductase [lm
10	295	23.3	248	1 B39930	Interneal B - Lis
11	294	23.2	248	2 AD1489	oxidoreductase hom
12	292	23.1	248	2 AD1128	oxidoreductase hom
13	289	22.8	247	2 AG1672	3-ketoacyl-acyl-ca
14	288.5	22.8	252	2 B82181	oxidoreductase, sh
15	288	22.8	248	2 T44932	3-oxoacyl-acyl-ca
16	287.5	22.7	246	2 C83961	3-oxoacyl-acyl-ca
17	286	22.6	247	2 F84210	oxidoreductase hom
18	285	22.5	247	2 AG1300	3-ketoacyl-acyl-ca
19	284	22.5	246	2 A69621	3-oxoacyl-acyl-ca
20	283	22.4	246	2 H72219	3-oxoacyl-acyl-ca
21	282	22.3	284	2 T02257	probable short cha
22	281	22.2	233	2 T35242	hypothetical prote
23	280	22.1	255	2 AF2466	NAD/NADP dependent
24	279.5	22.0	286	2 AI3207	hypothetical prote
25	278.5	22.0	244	2 S76993	3-oxoacyl-acyl-ca
26	278	22.0	246	2 H84136	oxidoreductase xfl
27	278	22.0	247	2 F82643	probable 2-hydroxy
28	276	21.9	255	2 T51764	3-oxoacyl-acyl-ca
29	277	21.9	246	2 A87457	

30	277	21.9	251	2 AH2042	3-oxoacyl-acyl-ca
31	275.5	21.8	253	2 B86737	acetoin dehydrogen
32	273.5	21.6	260	2 H70758	probable fabG3 pro
33	272.5	21.5	272	2 A99950	hypothetical prote
34	272	21.5	266	2 F83127	probable short-cha
35	271.5	21.5	260	2 G84171	oxidoreductase [lm
36	271	21.4	238	1 D69930	probable 3-oxoacyl
37	270	21.3	249	2 B97605	cyclohexanol dehyd
38	270	21.3	249	2 AD2827	short chain dehydr
39	269	21.3	257	2 AB1866	hypothetical prote
40	268.5	21.2	248	2 F82128	3-oxoacyl-acyl-ca
41	268.5	21.2	248	2 F81971	probable 3-oxoacyl
42	268.5	21.2	252	2 H64122	ydfG protein - Hae
43	267	21.1	258	2 H75616	oxidoreductase, sh
44	267	21.1	278	2 E83152	probable short-cha
45	266.5	21.1	244	2 T12051	3-oxoacyl-acyl-ca

ALIGNMENTS

RESULT 1

S07134 ribitol 2-dehydrogenase (EC 1.1.1.56) - Klebsiella pneumoniae

C:Species: Klebsiella pneumoniae

C:Date: 18-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 24-Oct-2000

C:Accession: S07134; S07135

R:Dothle, J.M.; Giglio, J.R.; Moore, C.B.; Taylor, S.S.; Hartley, B.S.

Biochem. J. 230, 569-578, 1985

A:Title: Ribitol dehydrogenase of Klebsiella aerogenes. Sequence and properties of w1

A:Reference number: S07134; MUID:86050423; PMID:3904726

A:Accession: S07134

A:Molecule type: protein

A:Residues: 1-249 <DOT>

A:Note: artifactual S-carboxymethyl cysteine was found at positions 30 and 126 after

R:Loviny, T.; Norton, P.M.; Hartley, B.S.

Biochem. J. 230, 579-585, 1985

A:Title: Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the structural ge

A:Reference number: S07135; MUID:86050424; PMID:2933028

A:Accession: S07135

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-145, 'SS', 148-211, 'N', 213-249 <LOW>

A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: oxidoreductase

F:15-191/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 87.0%; Score 1100; DB 2; Length 249;

Best Local Similarity 85.9%; Pred. No. 5.4e-84;

Matches 214; Conservative 21; Mismatches 14; Indels 0; Gaps 0;

QY	2	MNHNSVPSMNTPLNGKVAATGAASGIGLOCAKTLLDAGAVVILDRGDKLHKTVAELGE	61
DB	1	MKHSVSMNTSLSGKVAATGAASGIGLECAKTLLDAGAVVILDRGDKLHKTVAELGE	60
QY	62	NAVALQDLFNNQOVNMLADIETELAGGLDIFPANGAVYIGCPVAEGDPVMDRVLLNT	121
DB	61	NARALQVDLMQAOOVNMLQGLIQLGRDIFPANGAVYIGCPVAEGDPVMDRVLLNT	120
QY	122	NAAFRCRAVLPMLIAORSGDIIFTSSIAGVPIWEPIYTASKFAVOAFVHTTRQVSQ	181
DB	121	NAAFRCRAVLPMLIAORSGDIIFTFAVAGVPIWEPIYTASKFAVOAFVHTTRQVAQ	180
QY	182	YGVRAVAVLPGRVVTALLDDMPRAKMEALANGSLMPITVAESYLEFMVTRSKNVTVRD	241
DB	181	YGVRAVAVLPGRVVTALLDDMPRAKMEALANGSLMPITVAESYLEFMVTRSKNVTVRD	240
QY	242	VILPGSVDL 250	
DB	241	VILPNSVDL 249	

RESULT 2

DEKAR
 ribitol 2-dehydrogenase (EC 1.1.1.56) - Enterobacter aerogenes
 C:Species: Enterobacter aerogenes
 C>Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 23-Mar-1995
 C:Accession: A94585; A00344
 R:Hartley, B.S.
 Submitted to the Atlas, June 1977
 A:Reference number: A94585
 A:Accession: A94585
 A:Molecule type: protein
 A:Residues: 1-247 <NR>
 A>Note: this enzyme binds the coenzyme nicotinic adenine dinucleotide
 R:Moore, C.H.; Taylor, S.S.; Smith, M.J.; Hartley, B.S.
 unpublished results, cited by Morris, H.R., Williams, D.H., Midwinter, G.G., and Hartley
 A:Reference number: A94480
 A:Contents: annotation
 A:Note: Cys-30 and Cys-126 are carboxymethylcysteine
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: oxidoreductase
 F:15-189/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 84.8%; Score 1073; DB 1; Length 247;
 Best Local Similarity 84.3%; Pred. No. 9.3e-82;
 Matches 210; Conservative 23; Mismatches 14; Indels 2; Gaps 1;

QY 2 MNHSVPMNTPLNCKVAITGASGIGLQCAKTLIDAGAKVLLIDREGDKLHKVAELGE 61
 1 MKHSVSMNTPSLCKVAITGASGIGLQCAKTLIDAGAKVLLIDREGDKLHKVAELGE 60
 DB 62 NAVLQDLDFNNOQVDMNLADIELAGLDIFHNAAGVYIGPVAEGDPDVMNLNNT 121
 61 NAVLQDLDFNNOQVDMNLADIELAGLDIFHNAAGVYIGPVAEGDPDVMNLNNT 120
 QY 122 NAEFCVRAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSO 181
 121 NAEFCVRAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSO 178
 DB 121 NAEFCVRAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSO 178
 QY 182 YGVAVGAVLPGPVYVTTALLDMPKAKMEBALANGSLMOPIEVAESVLFMTVRSKNVTVRD 241
 179 YGVAVGAVLPGPVYVTTALLDMPKAKMEBALANGSLMOPIEVAESVLFMTVRSKNVTVRD 238
 DB 242 VILPGSYDL 250
 239 VILPGSYDL 247

RESULT 3

AC3632
 ribitol 2-dehydrogenase (EC 1.1.1.56) [Imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AC3632
 R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AC3632
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-242 <KUR>
 A:Cross-references: GB:AE008918; PIDN:AA154222.1; PID:g17985193; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI10980
 A:Map position: II
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 66.6%; Score 843; DB 2; Length 242;
 Best Local Similarity 66.9%; Pred. No. 1.1e-62;
 Matches 162; Conservative 36; Mismatches 44; Indels 0; Gaps 0;

QY 9 MNPFLNGKVAITGASGIGLQCAKTLIDAGAKVLLIDREGDKLHKVAELGENAVLQL 68
 1 MAOPLNGKVAITGASGIGLQCAKTLIDAGAKVLLIDREGDKLHKVAELGENAVLQL 60
 DB 69 DLFNNOQVDMNLADIELAGLDIFHNAAGVYIGPVAEGDPDVMNLNNTAARCV 128
 61 DLDPASVATMPOILQAGLDIFHNAAGVYIGPVAEGDPDVMNLNNTAARCV 120
 DB 129 RAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSOYGVAVGA 188
 121 RAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSOYGVAVGA 180
 DB 189 VLPSPVYVTTALLDMPKAKMEBALANGSLMOPIEVAESVLFMTVRSKNVTVRDVILPGSV 248
 181 VLPSPVYVTTALLDMPKAKMEBALANGSLMOPIEVAESVLFMTVRSKNVTVRDVILPGSV 240
 QY 249 DL 250
 241 DL 242

RESULT 4

B98199
 ribitol 2-dehydrogenase (EC 1.1.1.56) [Imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 C:Accession: B98199
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
 A: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Marrelz,
 Science 294, 2323-2328, 2001
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: B98199
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-243 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89116.1; PID:g15158922; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_1076
 A:Map position: linear chromosome
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 45.3%; Score 573; DB 2; Length 243;
 Best Local Similarity 46.1%; Pred. No. 3.1e-40;
 Matches 111; Conservative 52; Mismatches 78; Indels 0; Gaps 0;

QY 9 MNPFLNGKVAITGASGIGLQCAKTLIDAGAKVLLIDREGDKLHKVAELGENAVLQL 68
 1 MNPFLNGKVAITGASGIGLQCAKTLIDAGAKVLLIDREGDKLHKVAELGENAVLQL 60
 DB 69 DLFNNOQVDMNLADIELAGLDIFHNAAGVYIGPVAEGDPDVMNLNNTAARCV 128
 61 NLDADSCNMATPILDKVHDILYCNAGYIGDLETTTPRALDKMLNNTAVNAVKV 120
 QY 129 RAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSOYGVAVGA 188
 121 RAVLPNMAIORSGLDIFTSIAGVPIWEPIYASKFAVAQAFVHTTRQVSOYGVAVGA 180
 DB 189 VLPSPVYVTTALLDMPKAKMEBALANGSLMOPIEVAESVLFMTVRSKNVTVRDVILPGSV 248
 181 VSPSPVYVTTALLDMPKAKMEBALANGSLMOPIEVAESVLFMTVRSKNVTVRDVILPGSV 240
 QY 249 D 249
 241 D 241

RESULT 5

AG3087
 ribitol 2-dehydrogenase rdb [Imported] - Agrobacterium tumefaciens (strain C58, Dupon
 C:Species: Agrobacterium tumefaciens
 C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002

A:Genome: plasmid
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match | 26.3%; Score 333; DB 2; Length 249;
Best Local Similarity 34.4%; Pred. No. 2,7e-20;
Matches 84; Conservative 45; Mismatches 105; Indels 10; Gaps 4;

OY 11 TPLNGKVAITGAASGIGLOCAKTLLDAGAKVYLIDREGKLH---KIYAELGENAYAL 66
| :
DB 2 TSLNKGIALTVYGASSGICAAATAAKLAEGKAVGIAARTRTKLEDLKKIKIANGGE-ALVI 60
:::
OY 67 OLDLFNNQOVNMADIETLAGLDIFHANAGAVIGSPVNEGPDVDYRLNINNAFR 126
DB 61 EMDVDVTTSVAGKKLVADAYGSDILVNNAGLMPLPSDDIOGFVKDEQRQWADVNVGCLLN 120
OY 127 CVRAVLPHMIQRSGLDIETSSIAGVPVPIWEPIYTASKFAVOAFVTTTTRQVSQ-YGVR 185
DB 121 TTAAVLPGMIMQHSGHYFNMSSIAGRKYFGKLSYYCATKAHVTAFSGLMEVGOKHGIR 180
OY 186 VGAVLPGEVVTALLDMDP---KAKMEBALANGSLMPIEVAESVLEPMTRSKNVTVDRL 241
DB 181 VTCIOPGAVATELYDHITDPGYROOMDELATQMFTLOGEDIGDTIVFAQAAPAHVDVAEL 240
OY 242 VILP 245
DB 241 FVLP 244

RESULT 7
G90050
hypothetical protein SA2266 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence-revision 10-May-2001 #text-change 22-Oct-2001
C:Accession: G90050
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, M., A.; Mizutani-Iii, Y.; Chikahara, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimura, A.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuyu, K.
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: AB9758; PMID:21311952; PMID:11418146
A:Accession: G90050
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <KUR>
A:Cross-references: GB:BA000018; PID:g13702428; PIDN:BAB43569.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2266
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match | 26.2%; Score 331; DB 2; Length 231;
Best Local Similarity 31.8%; Pred. No. 3.6e-20;
Matches 75; Conservative 54; Mismatches 99; Indels 8; Gaps 2;

OY 11 TPLNGKVAITGAASGIGLOCAKTLLDAGAKVYLIDREGKLKIYAELEGENAY-ALQLD 69
| :
DB 2 TVLTDKIAYVVVGAGSGIGEAATATLHEBGAKVYVLAGRKRKTLQNVANQLADSVKVPPTD 61
OY 70 LENNOOVNMADIETLAGLDIFHANAGAVIGSPVNEGPDVDYRLNINNAFCVR 129
DB 62 VTNNEVEVELMKIAQQTFGGGLDIVINSAGOMLSKTIYDOVEDMSIDVINIGTLYTAQ 121
OY 130 AVLPMMIQORSGLDIETSSIAGVPVPIWEPIYTASKFAVOAFVHTTRQVSQKGVRYGAV 189
DB 122 AALPMLEQSSGHILNIASISIGEFTVSTSYTSKFAAVHTITOGLEKEIAKTGVKVTST 181
OY 190 LPEPVYVALLDMPKAKMEBALANGSLMPIEVAESVLEPMTRSKNVTVDVILP 245
DB 182 SPGWAVDTAITAANYNPDSRRK-----LDPQDAIEAVLVALTLQPKHNVNMEITVPR 230

RESULT 8
I17307

Db 182 VTTPPAINTELLETTIDKETEGMT--SLYKQIGITPDRIASIVAYALDQPEDVNNNEF 239

A, Experimental source: serogroup O1; strain N16361; biotype E1 for

A, Experimental source: serogroup O1; strain N16361; biotype E1 for

C:Genetics:
A:Gene: VC1591
A:Map position: 1
C:Superfamily: rhlitol dehydrogenase; short-chain alcohol dehydrogenase homology

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:42:27 ; Search time 4.96689 Seconds
(without alignments)
2087.642 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265

Sequence: 1 MMNHSVPSMNTPLNGKVAAL.....TRSKNVTVRDVILPGSYDL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	87.4	249	1 RIDH_KLEAE	P00335 klebsiella
2	294	23.2	248	1 Y452_LISTIN	Q926K7 listeria in
3	292	23.1	248	1 Y432_LISTMO	P25145 listeria mo
4	284	22.5	246	1 FABG_BACSU	P51831 bacillus su
5	283	22.4	246	1 FABG_THEMA	Q9X248 thermotoga
6	280.5	22.2	250	1 LINX_PSEPA	P50198 pseudomonas
7	273.5	21.6	260	1 YK02_MYCTU	Q10855 mycobacteri
8	271	21.4	238	1 YOXD_BACSU	P14802 bacillus su
9	268.5	21.2	244	1 FABG_VIBCH	Q9K9H7 vibrio chol
10	266.5	21.1	252	1 YDFG_HAEIN	P45300 haemophilus
11	266.5	21.1	244	1 FABG_VIBHA	P55336 vibrio harv
12	264.5	20.8	261	1 DHBA_BACSU	P39071 bacillus su
13	262.5	20.8	248	1 FABG_AQUAE	Q67610 aquifex aeo
14	257	20.3	247	1 FAGI_SYNY3	P73374 synechocyst
15	253	20.0	267	1 HDHA_CLOSO	P50200 clostridium
16	251	19.8	263	1 UCFA_SALTY	P37441 salmonella
17	250.5	19.8	248	1 YDRG_ECOLI	P39831 escherichia
18	250.5	19.8	266	1 PGDH_HUMAN	P15428 homo sapien
19	246.5	19.5	258	1 BDHA_RHIME	O86034 rhizobium m
20	245.5	19.4	261	1 ACT3_STRCO	P15644 streptomyces
21	243.5	19.2	278	1 YALA_RHISM	P55541 rhizobium s
22	243	19.2	258	1 BDHA_ALCEU	Q9X622 alcaligenes
23	242.5	19.2	248	1 YDRG_SALTY	P40864 salmonella
24	241.5	19.1	244	1 FABG_SALTY	O85141 salmonella
25	241.5	19.1	320	1 FABG_CUPPLA	P28643 cuphea lanc
26	241	19.1	256	1 Y019_THEMA	Q56318 thermotoga
27	240	19.0	255	1 YWFD_BACSU	P39640 bacillus su
28	239.5	18.9	247	1 YD50_MYCTU	Q11020 mycobacteri
29	239.5	18.9	263	1 UCFA_ECOLI	P37440 escherichia
30	238.5	18.9	263	1 UCFA_ECOLI	O8XJ44 escherichia
31	237.5	18.8	248	1 FABG_CHLPP	Q928P2 chlamydia p
32	237.5	18.8	262	1 YXBG_BACSU	P46331 bacillus su
33	237	18.7	262	1 VER1_ASPPA	P50161 aspergillus

34	236	18.7	264	1 STCU_EMENT	Q00791 emerichia
35	231.5	18.3	255	1 HDHA_ECOLI	P25529 escherichia
36	231	18.3	267	1 YW71_YEAST	O05016 saccharomyc
37	230.5	18.2	244	1 FABG_ECOLI	P25716 escherichia
38	230.5	18.2	245	1 NODG_RHIS3	P72332 rhizobium s
39	230.5	18.2	256	1 DHSD_RHOSH	Q59787 rhodobacter
40	227.5	18.0	245	1 NODG_RHIME	P06234 rhizobium m
41	226.5	17.9	287	1 HETV_ANASP	P37694 anabaena sp
42	226	17.9	261	1 DHKR_STRCM	P41177 streptomyces
43	225.5	17.8	252	1 YC1K_ECOLI	P31808 escherichia
44	225	17.8	250	1 LINX_PSEPA	P50197 pseudomonas
45	224	17.7	248	1 PHAB_ACISP	P50203 acinetobact

ALIGNMENTS

RESULT 1	RIDH_KLEAE	STANDARD:	PRT:	249 AA.
ID	RIDH_KLEAE			
AC	P00335:			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Ribitol 2-dehydrogenase (EC 1.1.1.56) (RDH).			
GN	REDD.			
OS	Klebsiella aerogenes.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OX	Klebsiella.			
NCBI_TaxID=28451;				
RN	(1)			
RP	SEQUENCE.			
RC	STRAIN-A, and D:			
RX	MEDLINE=86050423; PubMed=3904726;			
RA	Dothe J.M., Giglio J.R., Moore C.H., Taylor S.S., Hartley B.S.;			
RT	"Ribitol dehydrogenase of Klebsiella aerogenes. Sequence and			
RT	properties of wild-type and mutant strains.";			
RL	Biochem. J. 230:569-578(1985).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86050424; PubMed=2933028;			
RA	Loviny T., Norton P.M., Hartley B.S.;			
RT	"Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the			
RT	structural gene.";			
RL	Biochem. J. 230:579-585(1985).			
RN	[3]			
RP	SEQUENCE OF 1-68 FROM N.A.			
RX	MEDLINE=85230591; PubMed=3891331;			
RA	Wu J., Anderson-Loviny T., Smith C.A., Hartley B.S.;			
RT	"Structure of wild-type and mutant repressors and of the control			
RT	region of the rib operon of Klebsiella aerogenes.";			
RL	EMBO J. 4:1339-1344(1985).			
CC	-1- CATALYTIC ACTIVITY: Ribitol + NAD(+) = D-ribulose + NADH.			
CC	-1- COFACTOR: BINDS THE COENZYME ADENINE DINUCLEOTIDE.			
CC	-1- SUBUNIT: HOMOTETRAMER.			
CC	-1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN A.			
CC	-1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES			
CC	(SDR) FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: M25606; AAA25140.1; -			
DR	EMBL: X02448; CAA26292.1; -			
DR	PIR: A00344; DEKBR.			
DR	PIR: S07134; S07134.			
DR	PIR: S07135; S07135.			
DR	HSSP: P47227; 1BDB.			

```

DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 20 43 NAD (BY SIMILARITY).
FT ACT_SITE 160 160 BY SIMILARITY.
FT MOD_RES 30 30 CARBOXYMETHYLATION.
FT MOD_RES 126 126 CARBOXYMETHYLATION.
FT VARIANT 196 196 A -> P (IN STRAIN-D).
FT CONFLICT 146 147 AV -> SS (IN REF. 2).
FT CONFLICT 212 212 N -> D (IN REF. 1).
SQ SEQUENCE 249 AA; 26514 MW; C34896C276CC8FEC CRC64;

Query Match 87.4%; Score 1105; DB 1; Length 249;
Best Local Similarity 86.3%; Pred. No. 2e-83;
Matches 215; Conservative 20; Mismatches 14; Indels 0; Gaps 0;

OY 2 MNHVPSPNPLNGKVAITGAAGIGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIGE 61
DB 1 MKHVSVMNNTSLSGKVAITGAAGIGLQCAKTLIDAGAKVLLIDREGDKLHKIVAEIGE 60
OY 62 NAVALQDLFENNQVDNMLADIIELAGLDFIHANAGAYIGCPVAEGDPDVMDEVLNINI 121
DB 61 NAFALQVDLMQADQVDNLGIIQLQTLGRDIFHMANAGAYIGCPVAEGDPDVMDEVLNINI 120
OY 122 NAARCRVAIVPHMIAQRSGDIIFTSSIAGVVPYIWEPIYASKFVAQAFVHTTRROYSQ 181
DB 121 NAARCRVAVSLPHLIAQRSGDIIFTAVIAGVVPYIWEPIYASKFVAQAFVHTTRROYAQ 180
OY 182 YGVAVGAVLPGPVYATALLDDPKAKMEBALANGSLMOPIEVASVLEFVTRSKVNTVYDL 241
DB 181 YGVAVGAVLPGPVYATALLDDPKAKMEBALANGSLMOPIEVASVLEFVTRSKVNTVYDI 240
OY 242 VILPGSYDL 250
DB 241 VILPNSVDL 249

RESULT 2
Y452_LISIN STANDARD; PRT; 248 AA.
ID Y452_LISIN
AC Q92EK7;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical oxidoreductase Lin0452 (EC 1.-.-.-).
GN LMO0452.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CLIP 11262 / Serovar 6a;
MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
BAguero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Eutlier K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maltounam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA Science 294:849-852(2001).
CC - SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
(SDR) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL596165; CAC95684.1; -.
DR ListList: LMO0452; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 9 33 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 248 AA; 26809 MW; EAEB01F110306F CRC64;

Query Match 23.2%; Score 294; DB 1; Length 248;
Best Local Similarity 31.6%; Pred. No. 4.1e-17;
Matches 77; Conservative 46; Mismatches 105; Indels 14; Gaps 5;

OY 13 LNKGVAAITGAASIGLQCAKTLIDAGAKVLLIDREGDKLHKIV---AEIGENAVLQL 68
DB 3 LKNVYIIITGASGIGETATALLAEKGAKLVAARVEKLEKIYQTIKASSGEAIFA-KT 61
OY 69 DLFENNQVDNMLADIIELAGLDFIHANAGAYIGCPVAEGDPDVMDEVLNINIAARCV 128
DB 62 DVTREDNKKLVLELAIEYGVVDALIFLWAGIMPSPUSALKEDMEQWIDINIGVLNGI 121
OY 129 RAVLPHMIAQRSGDIIFTSSIAGVVPYIWEPIYASKFVAQAFVHTTRROYSQYGV-VRV 186
DB 122 AAVLPFSIAQSGHIITATSSYAGLKAIVGAVGATKMAVNDLMEVLRMEASQGTNIRT 181
OY 187 GAVLPGPVYATALLDDPKAKMEBALAN---GSLMOPIEVASVLEFVTRSKVNTVYDL 241
DB 182 ATIVPAIINFLELTETIDTKETEGMTNLVYKQYGV-VTPDRIASIVAAVQDPEDINNEF 239
OY 242 VILP 245
DB 240 TVGP 243

RESULT 3
Y432_LISMO STANDARD; PRT; 248 AA.
ID Y432_LISMO
AC P25145;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical oxidoreductase Lmo0432 (EC 1.-.-.-) (ORFA).
GN LMO0432.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EGD / Serovar 1/2a;
MEDLINE=91292517; PubMed=1905979;
RA Galliard J.-L., Berche P., Frehel C., Gouin E., Cossart P.,
RT Entry of L. monocytogenes into cells is mediated by internalin, a
RT secreted protein reminiscent of surface antigens from Gram-positive
RT cocci".
RL Cell 65:1127-1141(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=EGD-e / Serovar 1/2a;
MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
BAguero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Eutlier K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maltounam A., Mata Vicente J., Ng E., Nedjari H.,

```

RA Morchison H. R., de Mendoza D., Cronan J.E. Jr.; in a cluster of
RT lipid biosynthesis genes." *J. Bacteriol.* 178:4794-4800(1996).
RL [2]
RN
RP
RC
R 1
R 2
R 3
R 4
R 5
R 6
R 7
R 8
R 9
R 10
R 11
R 12
R 13
R 14
R 15
R 16
R 17
R 18
R 19
R 20
R 21
R 22
R 23
R 24
R 25
R 26
R 27
R 28
R 29
R 30
R 31
R 32
R 33
R 34
R 35
R 36
R 37
R 38
R 39
R 40
R 41
R 42
R 43
R 44
R 45
R 46
R 47
R 48
R 49
R 50
R 51
R 52
R 53
R 54
R 55
R 56
R 57
R 58
R 59
R 60
R 61
R 62
R 63
R 64
R 65
R 66
R 67
R 68
R 69
R 70
R 71
R 72
R 73
R 74
R 75
R 76
R 77
R 78
R 79
R 80
R 81
R 82
R 83
R 84
R 85
R 86
R 87
R 88
R 89
R 90
R 91
R 92
R 93
R 94
R 95
R 96
R 97
R 98
R 99
R 100
R 101
R 102
R 103
R 104
R 105
R 106
R 107
R 108
R 109
R 110
R 111
R 112
R 113
R 114
R 115
R 116
R 117
R 118
R 119
R 120
R 121
R 122
R 123
R 124
R 125
R 126
R 127
R 128
R 129
R 130
R 131
R 132
R 133
R 134
R 135
R 136
R 137
R 138
R 139
R 140
R 141
R 142
R 143
R 144
R 145
R 146
R 147
R 148
R 149
R 150
R 151
R 152
R 153
R 154
R 155
R 156
R 157
R 158
R 159
R 160
R 161
R 162
R 163
R 164
R 165
R 166
R 167
R 168
R 169
R 170
R 171
R 172
R 173
R 174
R 175
R 176
R 177
R 178
R 179
R 180
R 181
R 182
R 183
R 184
R 185
R 186
R 187
R 188
R 189
R 190
R 191
R 192
R 193
R 194
R 195
R 196
R 197
R 198
R 199
R 200
R 201
R 202
R 203
R 204
R 205
R 206
R 207
R 208
R 209
R 210
R 211
R 212
R 213
R 214
R 215
R 216
R 217
R 218
R 219
R 220
R 221
R 222
R 223
R 224
R 225
R 226
R 227
R 228
R 229
R 230
R 231
R 232
R 233
R 234
R 235
R 236
R 237
R 238
R 239
R 240
R 241
R 242
R 243
R 244
R 245
R 246
R 247
R 248
R 249
R 250
R 251
R 252
R 253
R 254
R 255
R 256
R 257
R 258
R 259
R 260
R 261
R 262
R 263
R 264
R 265
R 266
R 267
R 268
R 269
R 270
R 271
R 272
R 273
R 274
R 275
R 276
R 277
R 278
R 279
R 280
R 281
R 282
R 283
R 284
R 285
R 286
R 287
R 288
R 289
R 290
R 291
R 292
R 293
R 294
R 295
R 296
R 297
R 298
R 299
R 300
R 301
R 302
R 303
R 304
R 305
R 306
R 307
R 308
R 309
R 310
R 311
R 312
R 313
R 314
R 315
R 316
R 317
R 318
R 319
R 320
R 321
R 322
R 323
R 324
R 325
R 326
R 327
R 328
R 329
R 330
R 331
R 332
R 333
R 334
R 335
R 336
R 337
R 338
R 339
R 340
R 341
R 342
R 343
R 344
R 345
R 346
R 347
R 348
R 349
R 350
R 351
R 352
R 353
R 354
R 355
R 356
R 357
R 358
R 359
R 360
R 361
R 362
R 363
R 364
R 365
R 366
R 367
R 368
R 369
R 370
R 371
R 372
R 373
R 374
R 375
R 376
R 377
R 378
R 379
R 380
R 381
R 382
R 383
R 384
R 385
R 386
R 387
R 388
R 389
R 390
R 391
R 392
R 393
R 394
R 395
R 396
R 397
R 398
R 399
R 400
R 401
R 402
R 403
R 404
R 405
R 406
R 407
R 408
R 409
R 410
R 411
R 412
R 413
R 414
R 415
R 416
R 417
R 418
R 419
R 420
R 421
R 422
R 423
R 424
R 425
R 426
R 427
R 428
R 429
R 430
R 431
R 432
R 433
R 434
R 435
R 436
R 437
R 438
R 439
R 440
R 441
R 442
R 443
R 444
R 445
R 446
R 447
R 448
R 449
R 450
R 451
R 452
R 453
R 454
R 455
R 456
R 457
R 458
R 459
R 460
R 461
R 462
R 463
R 464
R 465
R 466
R 467
R 468
R 469
R 470
R 471
R 472
R 473
R 474
R 475
R 476
R 477
R 478
R 479
R 480
R 481
R 482
R 483
R 484
R 485
R 486
R 487
R 488
R 489
R 490
R 491
R 492
R 493
R 494
R 495
R 496
R 497
R 498
R 499
R 500
R 501
R 502
R 503
R 504
R 505
R 506
R 507
R 508
R 509
R 510
R 511
R 512
R 513
R 514
R 515
R 516
R 517
R 518
R 519
R 520
R 521
R 522
R 523
R 524
R 525
R 526
R 527
R 528
R 529
R 530
R 531
R 532
R 533
R 534
R 535
R 536
R 537
R 538
R 539
R 540
R 541
R 542
R 543
R 544
R 545
R 546
R 547
R 548
R 549
R 550
R 551
R 552
R 553
R 554
R 555
R 556
R 557
R 558
R 559
R 560
R 561
R 562
R 563
R 564
R 565
R 566
R 567
R 568
R 569
R 570
R 571
R 572
R 573
R 574
R 575
R 576
R 577
R 578
R 579
R 580
R 581
R 582
R 583
R 584
R 585
R 586
R 587
R 588
R 589

```

DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; FALSE NEG.
KM Fatty acid biosynthesis; Oxidoreductase: NADP; Complete proteome.
FT ACT_SITE 9 33 NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
FT CONFLICT 23 23 D -> A (IN REF. 1).
SQ SEQUENCE 246 AA; 26282 MW; C6A391167D3237DC CRC64;

Query Match 22.5%; Score 284; DB 1; Length 246;
Best Local Similarity 31.8%; Pred. No. 2.7e-16;
Matches 78; Conservative 47; Mismatches 110; Indels 10; Gaps 5;

QY 13 LKGVAAITGAASGIGLQCAKTLIDAGAKVVLIDREGD--KLAKIVAE--LGENAYALQ 67
D 2 LKDTAITYTGAASRGISALDLAKSGANVY-VNYSGEAAKANEVDEIKSMGRKALAVK 60
QY 68 LDLENNQVDMNLADIIELAGGLDIFHANAGAYIGGVAEGDDPVMDRVNLNINAFFRC 127
D 61 ADVSNPEQVMKIKETLSVFSTIDILVNMAGITRDNLIRKKEDEMDVAININLKGVFNC 120
QY 128 VRAVLPHMIAQRSGDIIFTSSIAGVVPVIMEPIYTASKFAVQAFVHTTRQVSGYRVYG 187
D 121 TKAVTRQMKKRSRRIINVSIVSGNPGQANTVAAGVIGLTKSSAKELASRNITVN 180
QY 188 AVLPQPVYVALLDDMPKAKMEALANGSLM--OPIEVAESVLEPMVTR-SKNVTVRDLVI 243
D 181 AIAAGFIETDPTDKLADVQDEMLKQIPLARFGEPSDVSVVTFLASGAYVMGQTLHI 240
QY 244 LPSGV 248
D 241 DGMV 245

RESULT 5
FABG_THEME STANDARD: PRT; 246 AA.
AC Q9248;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-acyl carrier protein reductase).
GN FABG OR TM1724.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; Pubmed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";
RT genome sequence of Thermotoga maritima.;
RL Nature 399:323-329(1999).
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb.ch/announce/
CC or send an email to license@lsb.ch).
CC -----
DR EMBL: AE001811; AAD36790.1; -.
DR HSSP: P50162; IAE1.
DR TIGR: TM1724; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KM Fatty acid biosynthesis; Oxidoreductase: NADP; Complete proteome.
FT ACT_SITE 10 34 NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 246 AA; 26401 MW; BC08904D28099142 CRC64;

Query Match 22.4%; Score 283; DB 1; Length 246;
Best Local Similarity 30.7%; Pred. No. 3.2e-16;
Matches 75; Conservative 47; Mismatches 112; Indels 10; Gaps 4;

QY 13 LKGVAAITGAASGIGLQCAKTLIDA--GAKVVLIDREGDKLAKIVAE--LGENAYALQ 67
D 3 LKGVCLITGAASIGK--ATTLFADGATVINGDISKEVLSVLAEBGLPKVDPVY 60
QY 68 LDLENNQVDMNLADIIELAGGLDIFHANAGAYIGGVAEGDDPVMDRVNLNINAFFRC 127
D 61 LVNTRDQIKIEVKEVQKGRIDVLVNMAGITRDALLVRKKEDEMDVAININLKGVFNC 120
QY 128 VRAVLPHMIAQRSGDIIFTSSIAGVVPVIMEPIYTASKFAVQAFVHTTRQVSGYRVYG 187
D 121 TVNVVPMIRKNGSIVNVSSVVGIVGNPGQNTVAASKAGVIGTKTWAELAGRNIRVN 180
QY 188 AVLPQPVYVALLDDMPKAKMEALAN--GSLQPIEVAESVLEPMVTRSKNVTVRDLVIL 244
D 181 AVAGFIETPTETKPLKAEFTALSRIPLGRFPEVAQVILFLADESSYVYGVI 240
QY 245 PGV 248
D 241 DGGL 244

RESULT 6
LINK_PSEPA STANDARD: PRT; 250 AA.
AC P50198;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase (EC 1.1.-.-) (2,5-DDOL dehydrogenase).
GN LINK.
OS Pseudomonas paucimobillis (Sphingomonas paucimobillis).
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=13689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UT26;
RX MEDLINE=94252977; Pubmed=7515041;
RA Nagata Y., Ohtomo R., Miyauchi K., Fukuda M., Yano K., Takagi M.;
RT Cloning and sequencing of a 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase gene involved in the degradation of gamma-hexachlorocyclohexane in Pseudomonas paucimobillis.;
RL J. Bacteriol. 176:3117-3125(1994).
CC -1- FUNCTION: DEGRADATION OF 2,5-DICHLORO-2,5-CYCLOHEXADIENE-1,4-DIOL (2,5-DDOL) INTO 2,5-DICHLOROHYDROQUINONE (2,5-DCHQ). LINK IS NOT ESSENTIAL TO GAMMA-HCH DEGRADATION.
CC -1- PATHWAY: DEGRADATION OF GAMMA-HEXACHLOROCYCLOHEXANE.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D23722; BAA04939.1; -
 DR HSSP: P19992; 1HDC.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short.1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT.1.
 DR Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
 KM NP_BIND 9 NAD (BY SIMILARITY).
 FT ACT_SITE 156 156 BY SIMILARITY.
 FT ACT_SITE 156 156 BY SIMILARITY.
 SQ SEQUENCE 250 AA; 25492 MW; 8C52703FF76382CF CRC64;

Query Match 22.2%; Score 280.5; DB 1; Length 250;
 Best Local Similarity 31.0%; Pred. No. 5.3e-16;
 Matches 75; Conservative 47; Mismatches 113; Indels 7; Gaps 3;

OY 9 MNTPLNGKVAATGASGIGLCAKTLDDGAKVVLIDREGDKLHKIVAEI--GENAYA 65
 DB 1 MANRLAGKVALITGASGIGLCAKTLDDGAKVVLIDREGDKLHKIVAEI--GENAYA 60
 OY 66 LQDLDFNNOQVDMNLADIETELAGLDIFHANAGAYIGPVAEGDPDVMRLNINAAE 125
 DB 61 ILDDVTADASNNNAIAAVDGEGLTTLSTNAGIIHPGFEESIGMKMNAVNOTAIF 120
 OY 126 KCVRAVLPMLIAORSGLDIFTSSIAGVVPVIMPEITASKFAVAQAVHTTRQVSGYVR 185
 DB 121 LGIRAVKPKMEKRGSLINISSIEGLAGTVAAGHYATKFAVGLTKSTALEGPSGR 180
 OY 186 VCAVLPGPVVLTALDDMPKAKMEALAN--GSLMQPIEAEVSLPMVT-RSKNVTVRDL 241
 DB 181 VMTIVPGMNTPTTANVPDVLKQRTSQIPMGKLDGPDIDANGALFLASDEAKYITGVDL 240
 OY 242 VI 243
 DB 241 PI 242

RESULT 7
 YK02_MYCTU STANDARD; PRT; 260 AA.
 ID YK02_MYCTU
 AC Q10855;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative oxidoreductase RV2002 (EC 1.-.-.-).
 GN FABG3 OR RV2002 OR MT2058 OR MTCY39.16C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmler K., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Horsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z74025; CAA98414.1; -
 DR EMBL: AE007057; AAK46335.1; -
 DR HSSP: P19992; 1HDC.
 DR TIGR: MT2058; -
 DR Tuberculist: RV2002; -
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short.1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT.1.
 KM Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
 FT NP_BIND 11 35 NAD (BY SIMILARITY).
 FT ACT_SITE 153 153 BY SIMILARITY.
 FT CONFLICT 174 174 S -> G (IN REF. 2).
 SQ SEQUENCE 260 AA; 27030 MW; 0935A1ED36220B7 CRC64;

Query Match 21.6%; Score 273.5; DB 1; Length 260;
 Best Local Similarity 31.6%; Pred. No. 2.1e-15;
 Matches 77; Conservative 47; Mismatches 111; Indels 9; Gaps 4;

OY 9 MNTPLNGKVAATGASGIGLCAKTLDDGAKVVLIDREGDKLHKIVAEI--GENAYA 65
 DB 1 MGRLLIGKVALVSGARGKASHRVAVAGAKVYVEDIIDEBC---KAAAEALADARY 57
 OY 66 LQDLDFNNOQVDMNLADIETELAGLDIFHANAGAYIGPVAEGDPDVMRLNINAAE 125
 DB 58 VHLDDVTPAQWMTAAVDTAVAFGLHVLVNNAGILNIGTIEDYALTEWQRIIDVNLGVF 117
 OY 126 KCVRAVLPMLIAORSGLDIFTSSIAGVVPVIMPEITASKFAVAQAVHTTRQVSGYVR 185
 DB 118 LGIRAVKPKMEKRGSLINISSIEGLAGTVAAGHYATKFAVGLTKSTALEGPSGR 177
 OY 186 VCAVLPGPVVLTALDDMPKAKMEALANGSLMQPIEAEVSLPMVT-RSKNVTVRDLVIL 244
 DB 178 VNSIHGVLKTPMTDWPEDIFOTAL--GNAAPVEVSNLYVLADDESSYSGAEFYVD 235
 OY 245 PGSV 248
 DB 236 GGTV 239

RESULT 8
 YOXD_BACSU STANDARD; PRT; 238 AA.
 ID YOXD_BACSU
 AC P14802;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical oxidoreductase yoxd (EC 1.-.-.-).
 GN YOXD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN-168;
RX MEDLINE-91192601; PubMed-1849493;
RA Ann R.S., Wake R.G.;
RT "variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W33 chromosomes.";
RL Gene 98:107-112(1991).
RN [2]
RP SEQUENCE OF 62-238 FROM N.A.
RX STRAIN-168;
RX MEDLINE-88040469; PubMed-1118336;
RA Corrigan C.M., Haarsma J.A., Smith M.T., Wake R.G.;
RT "sequence features of the replication terminus of the Bacillus
RT subtilis chromosome.";
RL Nucleic Acids Res. 15:8501-8509(1987).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06168; CAA29533.1; -
CC EMBL: Z99114; CAB13743.1; -
CC PIR: S01270; S01270.
CC HSP: 070351; 1E3S.
CC Subtilist: Bg11048; yoxd.
CC Interpro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Hypothetical protein; Oxidoreductase; Complete proteome.
CC NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).
CC ACT_SITE 155 155 BY SIMILARITY.
CC SEQUENCE 238 AA; 25299 MW; 9CCTA8D1204DF248 CRC64;
SQ
Query Match 21.4%; Score 271; DB 1; Length 238;
Best Local Similarity 32.5%; Pred. No. 3e-15;
Matches 74; Conservative 37; Mismatches 101; Indels 16; Gaps 4;
QY 13 LNCNVAITGAASIGIGQCAKTLIDAGAKVYLIDREGDKLHKIYAE--IGENAYALQLD 69
DB 4 LQHTTALITGGSGIGATLALAKEVNIIGLIGRTSANEKVAEEKALGVKAAFAAD 63
QY 70 LFNNOQVDMNLADITIELAGLDIFHANAGAYIGGPAVEGDDVDVRLNINNAFCVR 129
DB 64 VKDADQVNOVAOYKEQLGIDIDILINNAGISKFGFLDSLADEMENTIIVNLMGVHYTR 123
QY 130 AVLPFHIAQRSGDITFTSSIAGVVPVIMEPIYTSKRFAYQAFVHTTRROYSOYGVRCAY 189
DB 124 AVLPEMERKAGDIINISSTAGCGAIVTSYASKFAVIGLTESLMQEVKKNIRVSAL 183
QY 190 LFGVYVALLDDWPKAKMEALANGS---LMQPIEVAESVLFMYTRSK 234
DB 184 TPSTVASDM-----SIELNLTDGNPEKVMQPDIAE---YVVAOLK 221

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-El Tor N16961 / Serotype O1;
RX MEDLINE-20406833; PubMed-10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -----
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF004276; AAF95169.1; ALT_INIT.
CC HSP: P18992; IHDC.
CC TIGR: VC2021; -.
CC Interpro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Fatty acid biosynthesis; Oxidoreductase; NADP. Complete proteome.
CC NP_BIND 10 34 NADP (BY SIMILARITY).
CC ACT_SITE 151 151 BY SIMILARITY.
CC SEQUENCE 244 AA; 25566 MW; 9FB2E8278D7CC3CE CRC64;
SQ
Query Match 21.2%; Score 268.5; DB 1; Length 244;
Best Local Similarity 32.3%; Pred. No. 4.9e-15;
Matches 72; Conservative 39; Mismatches 107; Indels 5; Gaps 2;
QY 13 LNCNVAITGAASIGIGQCAKTLIDAGAKVYLIDREGDKLHKIYAEIGENAYALQLD 72
DB 3 LEGKVALVTGASRGIGAKIAELLAEKRAKVIIGTSSGAAISDIYDGNKGKMAVNTN 62
QY 73 NQOVDMNLADITIELAGLDIFHANAGAYIGGPAVEGDDVDVRLNINNAFCVRAVL 132
DB 63 PESTIEAVLKAITTDEFGVDILVNNAGITRDNLARMKEEEMSDIMETNLTISIFRLSKAVL 122
QY 133 PHMTAQRSGDITFTSSIAGVVPVIMEPIYTSKRFAYQAFVHTTRROYSOYGVRCAYLPG 192
DB 123 RGMKKRKGRIITNVGSVVGITGNAGQANVAAKAGVIGFTKSMAREVASRGVTVNTVAPG 182
QY 193 PVVT-----ALDDWPKAKMEALANGSLMOPIEVAESVLFMYTR 231
DB 183 FIETDMTKALNDEQRTATIAQVPA-GRLDGDPREIATSAVAFLAS 224

```

```

RESULT 9
FABG_VTBCH
ID FABG_VTBCH STANDARD; PRT; 244 AA.
AC 09KOH7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-
DE acyl carrier protein reductase).
DE FABG OR VC2021.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;

```

```

RESULT 10
YDFG_HAEIN
ID YDFG_HAEIN STANDARD; PRT; 252 AA.
AC PA5200;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical oxidoreductase H11430 (EC 1.-.-.-).
DE H11430.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;

```

```

OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Felschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrichbeck T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. STRONG, TO OTHER BACTERIAL HOMOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32822; AAC23077.1; -.
DR HSSP: P50162; 1AE1.
DR TIGR: H11430; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
FT NP_BIND 7 31 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 150 150 BY SIMILARITY.
SQ SEQUENCE 252 AA; 27398 MW; 227FB28CE7D9A98 CRC64;

Query Match 21.2%; Score 268.5; DB 1; Length 252;
Best Local Similarity 31.8%; Pred. No. 5, 1e-15;
Matches 75; Conservative 41; Mismatches 109; Indels 11; Gaps 3;

OY 18 AAITGAASGIGLCATKTLIDAGAKVYLIDREGDKLHKYAELEGNAVALDLEFNQOVD 77
DB 6 ALVTGATGAGLAICKKLLIAGKVIIGTRRADRLAEIHSQIGNFLPLAFDIRDOAT 65
OY 78 NMLADIIEIAGGLDIFHANAGAVIG-GPYAEGDPDWDRLNLINAAFCVRAVLP 136
DB 66 NMLNTLPEGMQAVDLVNNAGLALGLEPAHKADLDQWYOMIDINIGLVTITLVLVPMW 125
OY 137 AQRSGDIIFTSSIAGVVPVIMPIYASKFAVOAFVHTTRROYSOYGVAVLP 192
DB 126 ARNYGQIIMLSIAGTYPYAGSNVYGTAKFVYQFSLNLRADLAGKIRVSNVEPGCG 185
OY 193 ---PVYTTALDDMPKAKMEALANGSLMOPIVEAESVLEPMVTRSKVTVYRDVILP 245
DB 186 TEFNSVNRHFGRDDEKRAKYE---NWSVOPEDIANIVLWLEHOOPEHVNIINRIEVM 238

RESULT 11
FABG_VIBHA STANDARD: PRT; 244 AA.
AC P55336;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-
DE acyl carrier protein reductase).
FABG.
GN Vibrio harveyi.
OS

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B392;
RX MEDLINE=96134997; PubMed=8550484;
RA Shen Z., Byers D.M.;
RT "Isolation of Vibrio harveyi acyl carrier protein and the fabg, acpP,
RT and fabf genes involved in fatty acid biosynthesis.";
RL J. Bacteriol. 178:571-573(1996).
CC -1 CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NAD(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1 PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39441; AAC43589.1; -.
DR HSSP: P19992; 1HDC.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
KW Fatty acid biosynthesis; Oxidoreductase; NADP.
FT NP_BIND 10 34 NADP (BY SIMILARITY).
FT ACT_SITE 151 151 BY SIMILARITY.
SQ SEQUENCE 244 AA; 25519 MW; FC41A165B8CDA9A CRC64;

Query Match 21.1%; Score 266.5; DB 1; Length 244;
Best Local Similarity 32.3%; Pred. No. 7, 2e-15;
Matches 72; Conservative 39; Mismatches 107; Indels 5; Gaps 2;

OY 13 LNSKVAITGAASGIGLCATKTLIDAGAKVYLIDREGDKLHKYAELEGNAVALDLEFN 72
DB 3 LKGIKALVTCASGICRAIALLVETGATVIGTATSEGAALISEYLEGNGKALALNVT 62
OY 73 NOOVDMNLADIIEIAGGLDIFHANAGAVIGPYAEGDPDWDVNLINAAFCVRAVLP 132
DB 63 VESIEATLKTINDECAIDIDLNNAGITRDNLIRMKDKDEMDINDINLTPIYRMSRAVL 122
OY 133 PHMIAQRSGDIIFTSSIAGVVPVIMPIYASKFAVOAFVHTTRROYSOYGVAVLP 192
DB 123 RGMKKRRAGRIIVGVSVMGNAQTNYAAAGVIGFTKSMAREVAASGVTYNTVAPG 182
OY 193 PVYTTALDDMPKAKMEALANGSLMOPIVEAESVLEPMV 231
DB 183 FIETDMTKALINDOQRATLSNVPA-GRLGDPREIASAVVELAS 224

RESULT 12
DHBA_BACSU STANDARD: PRT; 261 AA.
AC P39071;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EC 1.3.1.28)
DE (Cold shock protein CS14).
OS DHBA OR ENTA.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;

```

```

RX MEDLINE-97080557; PubMed-8921902;
RA Rowland B.M., Grossman T.H., Osburne M.S., Taber H.W.;
RT "Sequence and genetic organization of a Bacillus subtilis operon
RL encoding 2,3-dihydroxybenzoate biosynthetic enzymes.";
RN Gene 178:119-123(1996).
RN [2]
RP SEQUENCE OF 59-255 FROM N.A.
RC STRAIN-168:
RA MEDLINE-94040785; PubMed-8224884;
RX Adams R., Schumann W.;
RT "Cloning and mapping of the Bacillus subtilis locus homologous to
RT Escherichia coli ent genes.";
RN Gene 133:119-121(1993).
RN [3]
RP SEQUENCE OF 1-11.
RC STRAIN-168 / JH642;
RA Graumann P.L., Schmid R., Marahel M.A.;
RL Submitted (OCT-1997) to the SWISS-PROT data bank.
CC -1- CATALYTIC ACTIVITY: 2,3-dihydro-2,3-dihydroxybenzoate + NAD(+) =
CC 2,3-dihydroxybenzoate + NADH.
CC -1- PATHWAY: 2,3-dihydroxybenzoate biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: IN RESPONSE TO LOW TEMPERATURE.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U26444; AAC44630.1; -
DR EMBL: L08644; AA16899.2; -
DR EMBL: 299120; CAB5190.1; -
DR PIR: P06883; P06883.
DR HSSP: P08074; 1CYD.
DR Subtilisin: BG11019; dbpa.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR Oxidoreductase: NAD; Complete proteome.
DR NP_BIND: 12 36 NAD (BY SIMILARITY).
DR ACT_SITE: 157 157 BY SIMILARITY.
DR FT ACT_SITE 146 146 P->D (IN REF. 2).
DR FT CONFLICT 231 234 IADA->MRC (IN REF. 2).
DR FT CONFLICT 247 256 TMHNLGVDS->RCIFMRCAT (IN REF. 2).
SQ SEQUENCE 261 AA; 27494 MW; 00B0E8BA53AB407C CRC64;

Query Match 20.9%; Score 264.5; DB 1; Length 261;
Best Local Similarity 28.6%; Pred. No. 1,1e-14;
Matches 72; Conservative 48; Mismatches 115; Indels 17; Gaps 4;

QY 13 LNCNKAITTAAGSIGLOCAKTLTDAGAKVLLDREDGKLHKIYAEI---GENAYALQD 69
DB 6 IEGLIAITTAAGSIGCAVAVARTLASGAIHAANDYNEKLEKYSLSKAERHAEAPAD 65
QY 70 LFNNOQVDMNLADIETLACGLDIFHANAGAYIGGPAEGDPVDVRYLNLINAAFCVR 129
DB 66 VRDSAAIDETARIEREMGPIDLIVNVAAGVLRPGLIHSLSDEWEATFVSSTGVFNASR 125
QY 130 AVLPHMAQSSGDIIFTSIAGVVPVIMEPIYTSKFAVOAFVHTTRKQVSQIVRGAV 189
DB 126 SVSKYMDRRSGSLVTVGSNPAGVPRTSMAYASSKAAAVMTKCLGLEIAEYIRCNIV 185
QY 190 LPGVYVAL-LDMPKAKMEALANGS-----LMQIEVAESVLEFMT-RSKN 235
DB 186 SPGSTETDMQSLMADENGAEOYIKGSLTEFFKIGIPLKTLAKPSDIADAVLETVSGAGH 245
QY 236 VIVRVLIILPGS 247
: : : : :

```

```

DB 246 ITMHNLCVDGCA 257
RESULT 13
FABG_AOUAE
ID FABG_AOUAE STANDARD; PRT; 248 AA.
AC 067610;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-oxoacyl-[acyl-carrier protein] reductase (Ec 1.1.1.100) (3-ketoacyl-
DE acyl carrier protein reductase).
GN FABG OR AO.1716.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE-98196666; PubMed-9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000752; AAC07575.1; -
DR HSSP: P50163; 2AE1.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR Fatty acid biosynthesis: Oxidoreductase: NADP; Complete proteome.
DR NP_BIND: 12 36 NADP (BY SIMILARITY).
DR ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 248 AA; 26867 MW; 5CFD9EB9ADB3F2C5 CRC64;

Query Match 20.8%; Score 262.5; DB 1; Length 248;
Best Local Similarity 26.6%; Pred. No. 1,6e-14;
Matches 67; Conservative 56; Mismatches 112; Indels 17; Gaps 4;

QY 9 MNTPLNKVAITTAAGSIGLOCAKTLTDAGAKVLLDREDGKLHKIYAEI---GENAY 64
DB 1 MEIKLGKVSIVTGSTGIGRAIAEKLASAGSTYIITGSEKRAKAAVEIANKYGVAKH 60
QY 65 ALOLDLFNNOQVDMNLADIETLACGLDIFHANAG-----AYIGGPAEGDPVDVRYLNL 119
DB 61 GVENMLLSEESINAFEEIYNLDGIDILVNNAGITDKLFLRNSLD-----WEELVKV 115
QY 120 NINAFRCVRAVLPHMAQSSGDIIFTSIAGVVPVIMEPIYTSKFAVOAFVHTTRKQV 179
DB 116 NLITGTFVLTONSLKMKIKRGRIVNIVSSVYGFIGNGVNNYSTTKGILGFTSLAKEL 175
QY 180 SQYGVRVGAVLPGPV---VTALLDDWPKAKMEALANGSLMQIEVAESVLEFMT-RSKN 236
DB 176 APRNVLVNAVAPGFIETDMNAVISEEIKQYKKEQIPIGRCGSPREVANVYFLCSELASY 235
: : : : :

```

Oy. 237 TVRDLVLPQSV 248
 Db 236 ITGEVTHVNGM 247

RESULT 14

FAG1_SYNY3 STANDARD: PRT; 247 AA.
 ID FAG1_SYNY3
 AC P73574;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 3-oxoacyl-[acyl-carrier protein] reductase 1 (EC 1.1.1.100) (3-oxoacyl-acyl carrier protein reductase 1).
 GN FAG1 OR SLR0886.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1 CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
 CC -1 PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: D90907; BAA17614.1; -
 DR HSP: P50162; IAE1.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KW Fatty acid biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 11 35 NADP (BY SIMILARITY).
 FT ACT_SITE 156 156 BY SIMILARITY.
 SQ SEQUENCE 247 AA; 25724 MW; 91EBF9409C777FE20 CRC64;

Query Match 20.3%; Score 257; DB 1; Length 247;
 Best Local Similarity 30.2%; Pred. No. 4, 4e-14;
 Matches 68; Conservative 46; Mismatches 107; Indels 4; Gaps 2;

Oy 11 TPUNGVAAITGAAGSIGLQCAKTLDDAGAKVYL-IDREGDKLHKIVAEI--GENAVAL 66
 Db 2 TALTNOVALVTCASRGKATLALATGMKVVYNAQSSSTADAVVAELIANGGALIV 61
 Oy 67 QLDLENNQOVDMNLADIIELAGGLDIFHANAGAYIGGPAEGSDPDVDRVLNINIAFR 126
 Db 62 QANVANADVDOLIKTTLPKFSRIDLVNNAAGITRTDLLRMKLEDMQAVIIDLNTGLVFL 121
 Oy 127 CVRAVLPHTIAQRSGDIIETSSIAGVVPVIMEIYTASKEFAVOAFHTTRROVSQGVAV 186
 Db 122 CTAAVSKMLKGRSGRIITTSVAGMMGNPGQANNSSAAAGVYGFYTKYAKELASRGVTV 181
 Oy 187 GAVLPGPVVTALDDMPKAKMEBALNGSLMOPIEVAESVLPVTV 231

Db 182 NAAVPGFIATDMTENLNAEPILOFIPLARYGOPEEVAAGTIRFLAT 226

RESULT 15

HDHA_CLOSO STANDARD: PRT; 267 AA.
 ID HDHA_CLOSO
 AC P50200;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 7-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) (Bile acid 7-dehydroxylase) (7-alpha-HSDH).
 OS Clostridium sordeum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OX NCBI_TaxID=1505;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.
 RX STRAIN=ATCC 9714;
 RA Coleman J.P., Hudson L.L., Adams M.J.;
 RT "Characterization and regulation of the NADP-linked 7 alpha-hydroxysteroid dehydrogenase gene from Clostridium sordeum.";
 RL J. Bacteriol. 176:4865-4874(1994).
 CC -1 CATALYTIC ACTIVITY: 3-alpha,7-alpha,12-alpha-trihydroxy-5-beta-cholanate + NADP(+) = 3-alpha,12-alpha-dihydroxy-7-oxo-5-beta-cholanate + NADPH.
 CC -1 SUBUNIT: HOMOTETRAMER (PROBABLE).
 CC -1 INDUCTION: BY BILE ACID. EXPRESSED DURING GROWTH-PHASE.
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: I12058; AAA3556.1; -
 DR HSP: P25529; IAH1.
 DR InterPro: IPR002198; ADH_short.
 DR Pfam: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KW Oxidoreductase; NADP; Bile acid catabolism.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 158 158 BY SIMILARITY.
 SQ SEQUENCE 267 AA; 29177 MW; 8560207B9EDCE0E CRC64;

Query Match 20.0%; Score 253; DB 1; Length 267;
 Best Local Similarity 31.2%; Pred. No. 1e-13;
 Matches 72; Conservative 37; Mismatches 106; Indels 16; Gaps 5;

Oy 13 LNKVAAITGAAGSIGLQCAKTLDDAGAKVYLIDREGDKLHKIVAEI--GENAVAL 72
 Db 4 LENKVALVTSATRGIGLASIKLQNGALVYKWRRLAEQTQCDYKKEGLTKPFVFD 63
 Oy 73 NQOVD--NMADIIELAGGLDIFHANAGAYIGP-----VAGDDPDVDRVLNINIA 124
 Db 64 ANVIDIYKKEIDITIKNECKIDILVNNFG-GRPEKDLVNGDEDTPEELFNVAVGSV 121
 Oy 125 FRCVAVLPHTIAQRSGDIIETSSIAGVVPVIMEIYTASKEFAVOAFHTTRROVSQGVAV 184
 Db 122 YRLSKLIIPIHMIENKGSIVNISVGSIPDISRIGVGSKSVNNITKOIAIYQAKYGI 181
 Oy 185 RVAVLPGPVVT-----ALLDDMPKAKMEBALNGSLMOPIEVAESVLPVTV 230
 Db 182 RCNAVLPGLIATDAANSMPEDEFRKSFVLSHVLN-RIGNEDIANSVLPFV 231

Tue Mar 25 12:01:56 2003

us-09-802-208b-3.rsp

Page 10

Search completed: March 13, 2003, 16:53:42
Job time : 6.96689 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:43:37 ; Search time 15.9354 Seconds
(without alignments)
3332:331 Million cell updates/sec

Title: US-09-802-208B-3

Perfect score: 1265
Sequence: 1 MMHNSVPSMNTPLNGKVAAL.....TRSKNVTYRDVILFSGVDL 250

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp_unclassified:*
- 15: sp-virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1265	100.0	250	Q9F4L7	Q9F4L7 escherichia
2	843	66.6	242	16 Q8YBC0	Q8YBC0 bruceella me
3	573	45.3	243	16 Q8U7X6	Q8U7X6 agrobacteri
4	366	28.9	247	2 Q9LCV7	Q9LCV7 streptomyce
5	342	27.0	247	2 Q54197	Q54197 streptomyce
6	333	26.3	249	16 Q930M2	Q930M2 rhizobium m
7	331	26.2	231	16 Q98RPF	Q98RPF staphylococ
8	317	25.1	248	2 Q9RH22	Q9RH22 zymomonas m
9	309	24.4	325	4 Q9BTF9	Q9BTF9 homo sapien
10	304	24.0	325	4 Q9UFW6	Q9UFW6 homo sapien
11	303.5	24.0	253	16 Q92NFB	Q92NFB rhizobium m
12	298	23.6	248	2 Q9LA07	Q9LA07 paenibacill
13	295.5	23.4	246	2 Q9EX74	Q9EX74 rhodococcus
14	295.5	23.4	255	1 Q34187	Q34187 halobacteri
15	292	23.1	248	2 Q9EXG3	Q9EXG3 listeria mo
16	291.5	23.0	258	16 Q982J8	Q982J8 rhizobium 1

17	289	22.8	247	16 Q92AK1	Q92AK1 listeria in
18	288.5	22.8	252	16 Q9KRP5	Q9KRP5 vibrio chol
19	288	22.8	248	2 Q44326	Q44326 agrobacteri
20	287.5	22.7	246	16 Q9KA03	Q9KA03 bacillus ha
21	287.5	22.7	257	16 Q92MR6	Q92MR6 rhizobium m
22	287.5	22.7	269	2 Q938D4	Q938D4 mycobacteri
23	286	22.6	247	17 Q9HRV8	Q9HRV8 halobacteri
24	285	22.5	247	16 Q8Y690	Q8Y690 listeria mo
25	283.5	22.4	256	16 Q982N5	Q982N5 rhizobium 1
26	282.5	22.3	245	16 Q8RDH9	Q8RDH9 thermomanae
27	282	22.3	260	5 Q15744	Q15744 dictyostell
28	282	22.3	284	10 Q50038	Q50038 nicotiana t
29	281.5	22.3	256	16 Q981H5	Q981H5 rhizobium 1
30	281.5	22.3	306	10 Q9LW34	Q9LW34 arabidopsis
31	281.5	22.3	306	10 Q941E4	Q941E4 arabidopsis
32	281	22.2	233	16 Q86732	Q86732 streptomyce
33	280.5	22.2	300	10 Q9LW35	Q9LW35 arabidopsis
34	280	22.1	255	16 Q8YL12	Q8YL12 anabaena sp
35	279.5	22.1	286	16 Q8UUT4	Q8UUT4 agrobacteri
36	278.5	22.0	244	16 Q55922	Q55922 synechocyst
37	278.5	22.0	291	4 Q9Y3A1	Q9Y3A1 homo sapien
38	278	22.0	246	16 Q9K636	Q9K636 bacillus ha
39	278	22.0	247	16 Q9PCN4	Q9PCN4 xylella fas
40	278	22.0	251	16 Q92PX8	Q92PX8 rhizobium m
41	278	22.0	255	2 Q07457	Q07457 rhodospseudo
42	277	21.9	246	16 Q9A7P5	Q9A7P5 caulobacter
43	277	21.9	251	16 Q8YV70	Q8YV70 anabaena sp
44	277	21.9	256	16 Q92MR3	Q92MR3 rhizobium m
45	276	21.8	254	16 Q988J4	Q988J4 rhizobium 1

ALIGNMENTS

RESULT 1

ID Q9F4L7 PRELIMINARY; PRT; 250 AA.

AC Q9F4L7;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Ribitol dehydrogenase.

GN R11D.

OS Escherichia coli;

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI_TaxID=562;

OX [1]

RN SEQUENCE FROM N.A.

RC STRAIN=C.

RA Lafayette P.R., Parrott W.A.

RT "A non-antibiotic marker for amplification of plant transformation vectors in E. coli."

RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.

CC EMBL: AY005817; AAC01883.1; -

DR HSSP: O70351; 1E6W.

DR InterPro: IPR002198; ADH_short.

DR Pfam: PF00106; adh_short; 1.

DR PRINTS: PR00080; SDRFAMILY.

DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.

KW Oxidoreductase.

SO SEQUENCE 250 AA; 26693 MW; 1787DACBFD6CDE CRC64;

Query Match 100.0%; Score 1265; DB 2; Length 250;
Best local Similarity 100.0%; Pred. No. 6e-91;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMHNSVPSMNTPLNGKVAALITGAASGIGLCARLTLDAGAKAVYLDESGKLRKYVAELG 60
|||||
DB 1 MMHNSVPSMNTPLNGKVAALITGAASGIGLCARLTLDAGAKAVYLIDREGDKLRKYVAELG 60

```
QY 61 ENAYALDLDLFNNQOVNMLADIETELAGLDIFHANAGATGCPVAEGDPDVMNRLN 120
DB 61 ENAYALDLDLFNNQOVNMLADIETELAGLDIFHANAGATGCPVAEGDPDVMNRLN 120
QY 121 INAFRCVAVLPHMIAQRSGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRQVS 180
DB 121 INAFRCVAVLPHMIAQRSGDIIFTSSIAGVPIWEPIYATASKFAVOAFVHTTRQVS 180
QY 181 QYGVAVGAVLPGPVYATLDDMPKAKMEALANGSLMOPIEVAESVLFMTTRSKNVTVRD 240
DB 181 QYGVAVGAVLPGPVYATLDDMPKAKMEALANGSLMOPIEVAESVLFMTTRSKNVTVRD 240
QY 241 LVILPGSVDL 250
DB 241 LVILPGSVDL 250

RESULT 2
O8YBCO PRELIMINARY; PRT; 242 AA.
ID O8YBCO;
AC O8YBCO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribitol 2-dehydrogenase (EC 1.1.1.56).
CN BMEI10980.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxId=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lyfidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selskov E., Elzer P.H., Hagius S., O'Callaghan D., Telesson J.-J.,
RA Heselkov R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009731; AAL54222.1; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adn_short.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 242 AA; 25972 MW; E41DB241DB237D1 CRC64;

Query Match 66.68; Score 843; DB 16; Length 242;
Best Local Similarity 66.94; Pred. No. 4.6e-58;
Matches 162; Conservative 36; Mismatches 44; Indels 0; Gaps 0;
```

```
RESULT 3
O8U7X6 PRELIMINARY; PRT; 243 AA.
ID O8U7X6;
AC O8U7X6;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribitol 2-dehydrogenase.
CN RDI OR ATU4323 OR AGR_L1076.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxId=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D., Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Shenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gätling S., Miller N., Blanchard M.,
RA Querrollo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009360; AAL45117.1; -.
DR EMBL: AE008252; AAK89116.1; -.
KW Complete proteome.
SQ SEQUENCE 243 AA; 26264 MW; AC363D61F1564DB2 CRC64;

Query Match 45.38; Score 573; DB 16; Length 243;
Best Local Similarity 46.18; Pred. No. 5.3e-37;
Matches 111; Conservative 52; Mismatches 78; Indels 0; Gaps 0;
```


RESULT 4

09LCV7 PRELIMINARY: PRT: 247. AA.

ID 09LCV7: 01-OCT-2000 (TREMblrel. 15, Created)

AC 09LCV7: 01-OCT-2000 (TREMblrel. 15, last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, last annotation update)

DE Clavaldenhyde dehydrogenase.

GN CAD.

OS Streptomyces clavuligerus.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_Taxid=1901;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL3585;

RX MEDLINE-94374706; PubMed-8088547;

RA Aldoo K.A., Wong A., Alexander D.C., Rittammer R.A., Jensen S.E.;

RT "Cloning, sequencing and disruption of a gene from Streptomyces clavuligerus involved in clavulanic acid biosynthesis.";

RL Gene 147:41-46(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL3585;

RX MEDLINE-97047987; PubMed-8892828;

RA Paraskar A.S., Aldoo K.A., Wong A., Jensen S.E.;

RT "Molecular analysis of a beta-lactam resistance gene encoded within the cephamycin gene cluster of Streptomyces clavuligerus.";

RL J. Bacteriol. 178:6266-6274(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL3585;

RX MEDLINE-98175459; PubMed-9515708;

RA Paraskar A.S., Aldoo K.A., Jensen S.E.;

RT "A pathway-specific transcriptional activator regulates late steps of clavulanic acid biosynthesis in Streptomyces clavuligerus.";

RL Mol. Microbiol. 27:831-843(1998).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL3585;

RX MEDLINE-20145396; PubMed-10681345;

RA Jensen S.E., Elder K.J., Aldoo K.A., Paraskar A.S.;

RT "Enzymes catalyzing the early steps of clavulanic acid biosynthesis are encoded by two sets of paralogous genes in Streptomyces clavuligerus.";

RL Antimicrob. Agents Chemother. 44:720-726(2000).

CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.

DR EMBL: U87786; AAF86624.1; -.

DR HSSP: P14061; 1FDS.

DR InterPro: IPR002198; ADH_short.

DR Pfam: PF00106; adh_short.1.

DR PRINTS: PR00080; SDRFAMILY.

KW Oxidoreductase.

SQ SEQUENCE 247 AA; 26325 MW; 30A135E8CF7E2E6D CRC64;

Query Match 28.9%; Score 366; DB 2; Length 247;

Best Local Similarity 37.3%; Pred. No. 7.5e-21;

Matches 91; Conservative 41; Mismatches 104; Indels 8; Gaps 3;

Db 180 VVIEPGETDTLGRHITHATATKEMEYQRISQIRKLDQADIAVAVYATAPRHATVHEI 239

QY 242 VILP 245

Db 240 FIRP 243

RESULT 5

054197 PRELIMINARY: PRT: 247. AA.

ID 054197: 01-JUN-1998 (TREMblrel. 06, Created)

AC 054197: 01-JUN-1998 (TREMblrel. 06, last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)

DE Clavulanate-9-aldehyde reductase.

GN CAR.

OS Streptomyces clavuligerus.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_Taxid=1901;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 27064;

RX MEDLINE-98267213; PubMed-9602162;

RA Perez-Redondo R., Rodriguez-Garcia A., Martin J.F., Liras P.;

RT "The clar gene of Streptomyces clavuligerus, encoding a LysR-type regulatory protein controlling clavulanic acid biosynthesis, is linked to the clavulanate-9-aldehyde reductase (car) gene.";

RL Gene 211:311-321(1998).

RL -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.

CC (SDR) FAMILY.

DR EMBL: AJ000671; CAA04227.1; -.

DR HSSP: P14061; 1FDY.

DR InterPro: IPR002198; ADH_short.

DR Pfam: PF00106; adh_short.1.

KW Oxidoreductase.

SQ SEQUENCE 247 AA; 26627 MW; 251C1E2FCE6E75619 CRC64;

Query Match 27.0%; Score 342; DB 2; Length 247;

Best Local Similarity 35.7%; Pred. No. 5.6e-19;

Matches 87; Conservative 40; Mismatches 109; Indels 8; Gaps 3;

QY 9 MNTPLNGKVAALITGAASGIGLQCAKTLTDAGAKVLLDREGDKLHKIVAE--GENAYA 65

Db 1 MPALOGKVALITGREGHRRATRALAPGCAVAITAARVEKLRALGDELTAAGAKVHV 60

QY 66 LQDLFNNQOVNMLADIIELAGLDIFHANAGAYIGPVAEGDPDVMRVNLINAAE 125

Db 61 LEIDVADRCGVDAAVASTVEALGDLILVNNAGIMLGPVEDADDTDMFTMIDTNILGLM 120

QY 126 RCYRAVLPHMIAQRSDIITSSIAGVVPIWEPITASKFAVQAVHTTRRVSQYGR 185

Db 121 YMTRALPHLRSK-GTVVQMSIAGRVVNRNAAYQATKFGVNAFSEVTRROEYTERGV 179

QY 186 VGAVLPGPVYAL---LDDWPKAKMEALANGSLMQPIEVAESVLEFMTRSKNVTVRDL 241

Db 240 FIRP 243

RESULT 6

0930M2 PRELIMINARY: PRT: 249. AA.

ID 0930M2: 01-DEC-2001 (TREMblrel. 19, Created)

AC 0930M2: 01-DEC-2001 (TREMblrel. 19, last sequence update)

DT 01-MAR-2002 (TREMblrel. 20, last annotation update)

DE Putative.

GN RA0173 OR SMA0320.

OC Rhizobiaceae; Sinorhizobium.
RX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gohier T., Goffeau A., Kahn D., Käss D., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ransperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021".
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591790; CAC46826.1; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 253 AA; 26818 MW; 0F2AAE2B89F4D8C3 CRC64;

Query Match 24.0%; Score 303.5; DB 16; Length 253;
Best Local Similarity 35.5%; Pred. No. 5.8e-16;
Matches 87; Conservative 43; Mismatches 100; Indels 15; Gaps 4;

QY 13 LNCVAAITGASGIGLOCAKTLIDAGAKVYLDR-----GDKLHKIVAELEGNAVALQL 68
DB 3 LNNRIATVTGAGSGIGRAGAIAIRGCAHVVYVDRSVEAGDFVAAIAG-GGSAEALAV 61

QY 69 DLFNNOGVNMLADIETELAGGLDIFHANAGAYIGGPAEGDPDWDVRLNINIAAFRCV 128
DB 62 DVYDDDLADGIADILYRHRGIDILHNHAGAVAGDIEVEVAGDFDSMNLNVAHFMMA 121

QY 129 RAVLPHMIAORSGDIETSSAGVVPVWEIYTSKFAVQAFVHTTRROYSGYGVGA 188
DB 122 RLVPSPKKKRGVYIVTSSSGVLXDREMIAYTTTHAVIYAMTROMAGDVAKKGVVNA 181

QY 189 VLPGPVYTFALLDW-----PKAKMEFALANGSLMQPLEVASEVLFPMWT-RSKNVTY 238
DB 182 LCPGWDVTPNEPPIDMGGRGAEIAYIRERVLGRMASVDEIAESILFLVSDRSYMTG 241

QY 239 RDLVI 243
DB 242 QILVV 246

RESULT 12
Q91A07 PRELIMINARY; PRT; 248 AA.
AC Q91A07;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 26.4 kDa protein.
OS Paenibacillus popilliae (Bacillus popilliae).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=78057;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14706;
RX MEDLINE=20145393; PubMed=10681342;
RA Patel R., Piper K.E., Cockerill F.R., Steckelberg J.M., Yousten A.A.;
RT "The biopesticide Paenibacillus popilliae has a vancomycin resistance
RT gene cluster homologous to the enterococcal VanA vancomycin resistance
RT gene cluster".
RL Antimicrob. Agents Chemother. 44:705-709(2000).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC EMBL: AF155139; AAF36807.1; -.
DR HSSP: P14061; 1FDS.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.

DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Hypothetical protein; Oxidoreductase.
SQ SEQUENCE 248 AA; 26432 MW; A122A9FF7799B10C CRC64;

Query Match 23.6%; Score 298; DB 2; Length 248;
Best Local Similarity 33.6%; Pred. No. 1.5e-15;
Matches 83; Conservative 43; Mismatches 105; Indels 16; Gaps 6;

QY 13 LNCVAAITGASGIGLOCAKTLIDAGAKVYLIDREDKHKIVAE-----GENAVVALQL 68
DB 3 IENKVVITGASSIGENTAKLAEKAKVYLGRREHLYKVELEKSGCGAAVAV-T 61

QY 69 DLFN---NQVDNMLADIETELAGGLDIFHANAGAYIGGPAEGDPDWDVRLNINIAAF 125
DB 62 DVVPPDSQDLVOLAKDTE--GGVDVIFLNAGLIMPSPLELTKDDEMSVGVNKGVL 118

QY 126 RCVAVLPHMIAORSGDIETSSAGVVPVWEIYTSKFAVQAFVHTTRROYSGY-- 183
DB 119 NGIAAVLPTFISQSGHITNSSVAGLAVPGCAVYGTAVANLNLEVLMEASQEGTN 178

QY 184 VRVGAVLPGPVYTFALLDWPKAKME--ALANGSLMQPLEVASEVLFPMYTRSKNVTVD 240
DB 179 IRTATITPAINTBLCTITTDKNISEGMTALYEGYGISPDVAVIAFIDQPEDTVNE 238

QY 241 LVLPGS 247
DB 239 FTIGPTS 245

RESULT 13
Q9EX74 PRELIMINARY; PRT; 246 AA.
AC Q9EX74;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE SDR-like enzyme.
GN MHA.
OS Rhodococcus erythropolis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC14;
RX MEDLINE=21091964; PubMed=11157238;
RA Van der Vlugt-Berghmans C.J.B., van der Werf M.J.;
RT "Genetic and biochemical characterization of a novel monoterpene
RT epsilon-lactone hydrolase from Rhodococcus erythropolis DC14";
RL Appl. Environ. Microbiol. 67:733-741(2001).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC EMBL: AJ292535; CAC17805.1; -.
DR HSSP: P50163; 2AEL.
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short; 1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
KM Oxidoreductase.
SQ SEQUENCE 246 AA; 25058 MW; 6DECB136FE9AA34A CRC64;

Query Match 23.4%; Score 295.5; DB 2; Length 246;
Best Local Similarity 32.9%; Pred. No. 2.3e-15;
Matches 81; Conservative 42; Mismatches 110; Indels 13; Gaps 6;

QY 9 MNPLNCKVAITGASGIGLOCAKTLIDAGAKVYLIDRSGDKHKIVAELEGNAVALQL 68
DB 1 MVALESGVAVVVTGGSGIGACVROLCALGASVVAIVDFNATLAKSEFGDRAVAVEV 60

QY 69 DLFNNOGVNMLADIETELAGGLDIFHANAGAYIGPV--AEGDP--DVMPRVNLINIAA 124
DB 61 DVARVEDAERKRVTAVAHFGGLDAVNNAG--GVGPKASVGDGTGFEEMRRVLDVNLDA 118

```

DT 01-JUN-2002 (TREMBLREL. 21, Last annotation update)
DE Hypothetical 26.8 kDa protein.
OS Listeria monocytogenes
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=EGD;
RA Hain T., Pashalidis P., Hudel M., Chakraborty T., Dommann E.;
RT "Nucleotide sequence of the internalin operon from Listeria
RT monocytogenes EGD.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBDP databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AJ012346; CAC20627.1; -.
DR HSSP; O70351; 1b6w.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KM Hypothetical protein; Oxidoreductase.
SQ SEQUENCE 248 AA; 26849 MW; 992CB93B9EC011D6 CRC64;

Query Match 23.1%; Score 292; DB 2; Length 248;
Best Local Similarity 32.1%; Pred. No. 4.4e-15;
Matches 79; Conservative 47; Mismatches 106; Indels 14; Gaps 5.

QY 13 LNGVAAITTGAAAGSGIGLCACAKTLIDGAKAVYLIDREGDKLHIV----AELGENAVALLQ 68
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 IKNVYIITTGASSSGIGKRTALLLAEKGAKIYLAARVELLETIVOTIRANSEALFA-KT 61

QY 69 DLFNNQVDNNMLADIIEIAGGLDIFHANAGAYIGSPVAEGDDVDVRYLNINNAFCRV 128
   : : : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 62 DVFKEEDNRKRLVELAEIERYGKVDIAFLNAGIMPNSPLSLKEDDEWQMIDINIKGLNGI 121

QY 129 RAVLPHMAQNSGDIIEFSSIAGVVPVIMEPIYTASKFAVQAFVHTTRQVSOYG--VRV 186
   : | | | | | | | | | | | | | | | | : : : : : | | :
Db 122 AAVLPSPFIAQKSGHIAIASSVAAGLAKVPGAGVAYGATKMAVRDLMVEYLRESAQEGNINFT 181

QY 187 GAVLPGPVVTALLDDWPKAKMEALANSL-----MPIEVASEVLYFWTRSKNVTRDL 241
   : : : : : | : | : : : | : : : : | : | : : : : | :
Db 182 VTPIPAINTLELETITTKETEDQGMT--SLYKQYGITPRIASIYAVAIIDQEDVDVNVNEF 239

QY 242 VILPGS 247
   : | |
Db 240 TVGPFS 245

Search completed: March 13, 2003, 16:55:09
Job time : 18.9354 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:37 ; Search time 43.7632 Seconds
(without alignments)
1625.927 Million cell updates/sec

Title: US-09-802-208B-4

Perfect score: 2811

Sequence: 1 MTIRKTVIGVDVSGSGSVRAG.....HHRRYEAVKQLOHFAKLIRD 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2811	100.0	534	22	AAE09780
2	1172	41.7	551	23	AAE24140
3	1171	41.7	551	22	AAAM40293
4	1005.5	35.8	548	22	ABB70032
5	672	23.9	246	22	AAAB93914
6	553.5	19.7	554	22	AAU39145
7	448.5	16.0	560	19	AAW53950
8	435.5	15.5	593	22	AAU39179
9	434	15.4	252	22	AAAB9394
10	340	12.1	518	23	ABP39224

11	330.5	11.8	521	22	AAAG2039
12	330	11.7	164	22	AAAM42079
13	320.5	11.4	505	23	ABP30486
14	320.5	11.4	518	23	ABP27721
15	317	11.3	504	23	ABBA47468
16	309	11.0	519	23	ABBS5539
17	306	10.9	484	22	AAU34790
18	306	10.9	498	23	ABBS4582
19	304.5	10.8	494	22	AAAB6826
20	303	10.8	501	22	AAU35273
21	302.5	10.8	503	22	AAU35487
22	298.5	10.6	497	23	ABBA4654
23	294.5	10.5	496	22	AAAG82146
24	293.5	10.4	497	23	ABP39566
25	289	10.3	517	22	AAU37244
26	288	10.2	509	22	AAU34300
27	281.5	10.0	508	23	ABP26834
28	279.5	9.9	509	22	AAAG29332
29	279.5	9.9	509	22	AAAB79335
30	277.5	9.9	502	23	ABP26833
31	276.5	9.8	497	20	AAW94843
32	276.5	9.8	502	22	AAU34832
33	276.5	9.8	502	22	AAU37838
34	273.5	9.7	502	22	AAU38067
35	272	9.7	538	22	ABBB6928
36	269	9.6	498	17	AAW60271
37	265.5	9.4	526	17	AAW06440
38	265.5	9.4	501	21	AAV70727
39	264.5	9.4	496	20	AAW94309
40	264.5	9.4	496	22	AAU33825
41	264.5	9.4	498	22	AAU36742
42	262.5	9.3	501	22	AAU38195
43	261	9.3	501	23	ABBS4851
44	259	9.2	514	23	ABP39502
45	255.5	9.1	499	22	AAAG82195

ALIGNMENTS

RESULT 1
AAE09780
ID AAE09780 standard; Protein: 534 AA.
XX
AC AAE09780;
XX
DT 29-NOV-2001 (first entry)
XX
DE Escherichia coli strain C rrl operon encoding ribitol kinase.
XX
XX Positive selection system: metabolise: arabitol; ribitol; mannitol;
KW transgenic cell; marker gene: ribitol dehydrogenase; ribitol kinase;
KW ribitol transporter; rrl operon.
XX
XX Escherichia coli C.
OS
XX
PN WO200166779-A2.
XX
PD 13-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US07474.
XX
PR 08-MAR-2000; 2000US-0188291.
PR 15-AUG-2000; 2000US-0255595.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Parrott W, Lafayette P, Kane P;
XX
DR WPI: 2001-565596/63.
DR N-PSDB; AAD16811.
XX
PT Positively selecting transformed cells comprising selectable marker

PT gene and desired gene, from a cell population by using marker compounds
 PT e.g., arabinol, ribitol which confer selective advantage on transformed
 PT cells -
 XX Claim 17: Page 37; 37pp: English.
 CC The present invention relates to a positive selection system that
 CC involves conferring to transferred cells the ability to metabolise
 CC arabinol, ribitol and/or mannitol. The positive selection method is
 CC used in positively selecting transgenic cells from a population of cells
 CC using the positive selection method, the presence of the gene of
 CC interest in the genetically transformed cells may be determined without
 CC the disadvantages associated with traditional negative selection
 CC systems. Positive selection of the transformed cells is achieved without
 CC directly damaging the neighbouring non-transformed cells. The
 CC transformed cells may be identified by simple visual means without the
 CC use of a separate assay to determine the presence of a marker gene. This
 CC technique also avoids the release of antibiotics or other dangerous
 CC genes into the environment. The present sequence is Escherichia coli
 CC strain C. ribitol kinase encoded by ribitol operon (rli operon). The
 CC operon also encodes ribitol dehydrogenase and ribitol transporter.
 CC
 XX
 SQ Sequence 534 AA;
 Query Match 100.0%; Score 2811; DB 22; Length 534;
 Best Local Similarity 100.0%; Pred. No. 7, 6e-270;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTIRKTVIGVNGSGSVAGIFDNLGSLSHATEKITTTTRSGSRVDSQEIOWAVCS 60
 DB 1 MTIRKTVIGVNGSGSVAGIFDNLGSLSHATEKITTTTRSGSRVDSQEIOWAVCS 60
 QY 61 IRNALTLADVCAQSVAGIGFDATCSLVVLDKNGDPLPVSPBGDAKONIIYVMDHRAEQ 120
 DB 61 IRNALTLADVCAQSVAGIGFDATCSLVVLDKNGDPLPVSPBGDAKONIIYVMDHRAEQ 120
 QY 121 ERINATTHPVLYNYGKISPEMERPKILMKENKPEIYERAGOFEDLADFLTRATGDL 180
 DB 121 ERINATTHPVLYNYGKISPEMERPKILMKENKPEIYERAGOFEDLADFLTRATGDL 180
 QY 181 RSVCTVCKMTWLAHENVMDPFRITGLAELADEDFIRIGHNIVSPPTCGNGLTAA 240
 DB 181 RSVCTVCKMTWLAHENVMDPFRITGLAELADEDFIRIGHNIVSPPTCGNGLTAA 240
 QY 241 AEMGLPFTPVAVGLIDAHAGICTVGEAGALNNLAVFESTSSCTMASTTSPSFV 300
 DB 241 AEMGLPFTPVAVGLIDAHAGICTVGEAGALNNLAVFESTSSCTMASTTSPSFV 300
 QY 301 GPYTSAMVPGMLVEGGOSAGAAIDOLDHPAVEAREMAORVNOPLPVMLADRI 360
 DB 301 GPYTSAMVPGMLVEGGOSAGAAIDOLDHPAVEAREMAORVNOPLPVMLADRI 360
 QY 361 TAQPSDAVALAKGLHVVPEFLGNRAPADPHARAVICGLGERDLDNLALYIGLC 420
 DB 361 TAQPSDAVALAKGLHVVPEFLGNRAPADPHARAVICGLGERDLDNLALYIGLC 420
 QY 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROILDTCIPYITTOCCPEVL 480
 DB 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROILDTCIPYITTOCCPEVL 480
 QY 481 GAVAGNIAPSVGEAMQOFTHYDKYYPQERYOSLHRRYEAYKOLQHTAKLLRD 534
 DB 481 GAVAGNIAPSVGEAMQOFTHYDKYYPQERYOSLHRRYEAYKOLQHTAKLLRD 534
 RESULT 2
 AAE24140
 ID AAE24140 standard; Protein; 551 AA.
 AC AAE24140;
 XX
 XX 23-SEP-2002 (first entry)
 XX

DE Human kinase (PKIN)-11 protein.
 XX
 XX Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;
 KW acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;
 KW asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;
 KW development; hepatitis; cardiovascular; hypertension; drug screening;
 KW myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;
 KW fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;
 KW hypercholesterolaemia; obesity; gene therapy; cytostatic; anti-HIV;
 KW neuroprotective; hepatotropic; hypotensive; cardiatic; nephrotropic;
 KW hyperlipidaemia; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN W0200233099-A2.
 XX
 PD 25-APR-2002.
 XX
 PF 20-OCT-2001; 2001WO-0547728.
 XX
 PR 20-OCT-2000; 2000US-242410P.
 PR 27-OCT-2000; 2000US-244068P.
 PR 03-NOV-2000; 2000US-245708P.
 PR 09-NOV-2000; 2000US-247672P.
 PR 16-NOV-2000; 2000US-249565P.
 PR 22-NOV-2000; 2000US-252730P.
 PR 01-DEC-2000; 2000US-250807P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Gururajan R, Baughn MR, Walla NK, Elliott VS, Xu Y, Arvizu C;
 PI Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;
 PI Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM;
 PI Lal PG, Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;
 PI Thangavelu K, Khan FA, Ison CH;
 XX
 DR WPI: 2002-454603/48.
 DR N-PSDB: AAD38854.
 XX
 PT New human kinase polypeptide, for diagnosing, preventing and treating
 PT cancer, immune system disorders, growth and development disorders,
 PT cardiovascular disorders and lipid disorders -
 XX
 PS Claim 1; Page 160-161; 210pp: English.
 XX
 CC The invention relates human kinases (PKIN) and their corresponding
 CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
 CC treating and preventing cancer, an immune system disorder (e.g.,
 CC acquired immune deficiency syndrome (AIDS), Addison's disease, allergy,
 CC asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders
 CC affecting growth and development (e.g., arteriosclerosis, cirrhosis,
 CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial
 CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty
 CC liver, Gaucher's disease, Niemann-Pick's disease, hypercholesterolaemia,
 CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
 CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
 CC condition or a disease associated with the expression of PKIN in a
 CC biological sample. A composition comprising PKIN or an agonist or
 CC antagonist of PKIN is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional PKIN.
 CC PKIN is useful in a number of drug screening techniques and to analyse
 CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
 CC knockin humanised animals, or transgenic animals to model human diseases,
 CC and in somatic or germ-line gene therapy. The present sequence is human
 CC PKIN protein.
 XX
 SQ Sequence 551 AA;
 Query Match 41.7%; Score 1172; DB 23; Length 551;
 Best Local Similarity 45.0%; Pred. No. 4, 3e-107;
 Matches 242; Conservative 96; Mismatches 182; Indels 18; Gaps 8;
 QY 8 IGVGVGSGSVAGIFDNLGSLSHATEKITTTTRSGSRVDSQEIOWAVCSIRNALTL 67


```

Db      13 VGVAVGTGSGVRAALVDQSGVLLAFADQPIKNWEPGFNHHHSESSDIMAACCVYTKKRYVG 72
Oy      68 ADVCAQSVAGIGFDPATCSLVVLDKNGDPLVPSPGDAKONIYMMHRAEQAEERINATH 127
Db      73 IDL--NQIRGLGFDATCSLVVLDKQFHPRLPVNOGDSHNRVIMWIDHRAVSQNRINETK 130
Oy      128 HPLVANYGKISPEMETPKILMLKENMPET--YERAGOFPLDPLFWRATGDLARSVCY 186
Db      131 HSLVLOVGVGVSMVEKQAPKILMLKENLREICWDKAGHFDPPLPSMKATGVARSLCSL 190
Oy      187 TCKWTWLAHENRMDPDTFRITIGLAELADEDFIRIGHNIIVSPGTCGNGLTQAQAAEMGILL 246
Db      191 VCKWTYSA--EKGMDDSFWMKMGIEDEFVADNYSKIGNQVLPFGASLNGLTLPPEARLDGILL 249
Oy      247 PGTPPVAVGLIDAHAGIGTVG--VEGGAL-----NMLAYVFGTSSCTMASTSPSFV 296
Db      250 PGIAVAVASLIDAHAGLGVIGADVGRHGLICEQOPTYSLRAVLCGTSSCHMGISKDPFV 309
Oy      297 PGVWGPYYSAMVPGLMLVEGGSAGAAIDQLDFHPAVEAREMAQRVNPPLVWLADR 356
Db      310 PGVWGPYFSAMVPGFMLNEGGSQVTKLIDHMVQGHAAPELDQKATARQSIYAYLNSH 369
Oy      357 I-LEKTAQPSDAVALAKGLHAVPEFLGNRPADPRPARAVYICGLGMRDIDLNLAYIAG 415
Db      370 LDLIKRAQF--VGFLTVDLHWMPDFHGNRSPLADLTCLKGWTGLKSQDLDDLAIILYAT 427
Oy      416 LCGIGYGLROIIDAQTAQGVVSKNIYISGAGOHPLVROLADPTGCPVITTCOCEPVLL 475
Db      428 VOAILGSTRITTEAMEAGHSISTLFLCGLSKNPLVONHADITGMPVVLVSQEVESVLV 487
Oy      476 GSAILGAVAGNIAPSVGEAMQOFTHYDKYYUPOERYQSLHNRREYAV--KOLQHTAKLL 532
Db      488 GAAVLGACASGDPRASVGEAMAKMSKVYFPRLDQKKYUQVFLKLVENQKETL 545

```

RESULT 3
 ID AAM40293
 AAM40293 standard; Protein: 551 AA.

AC AAM40293;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3438.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Dirmanc RT;
 XX
 DR MPI: 2001-442253/47.
 DR N-PSDB: AA159449.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 XX such as central nervous system injuries -

PS Example 5; SEQ ID NO 3438; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM44213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 551 AA;

Query Match 41.7%; Score 1171; DB 22; Length 551;
 Best Local Similarity 45.0%; Pred. No. 5,5e-107;
 Matches 242; Conservative 96; Mismatches 182; Indels 18; Gaps 8;

```

Oy      8 IGVVSGSVRAIGIFDLNGSLSHATEKITTTTRSGRVQSSQETIQAQSCIRNATLT 67
Db      13 VGVAVGTGSGVRAALVDQSGVLLAFADQPIKNWEPGFNHHHSESSDIMAACCVYTKKRYVG 72
Oy      68 ADVCAQSVAGIGFDPATCSLVVLDKNGDPLVPSPGDAKONIYMMHRAEQAEERINATH 127
Db      73 IDL--NQIRGLGFDATCSLVVLDKQFHPRLPVNOGDSHNRVIMWIDHRAVSQNRINETK 130
Oy      128 HPLVANYGKISPEMETPKILMLKENMPET--YERAGOFPLDPLFWRATGDLARSVCY 186
Db      131 HSLVLOVGVGVSMVEKQAPKILMLKENLREICWDKAGHFDPPLPSMKATGVARSLCSL 190
Oy      187 TCKWTWLAHENRMDPDTFRITIGLAELADEDFIRIGHNIIVSPGTCGNGLTQAQAAEMGILL 246
Db      191 VCKWTYSA--EKGMDDSFWMKMGIEDEFVADNYSKIGNQVLPFGASLNGLTLPPEARLDGILL 249
Oy      247 PGTPPVAVGLIDAHAGIGTVG--VEGGAL-----NMLAYVFGTSSCTMASTSPSFV 296
Db      250 PGIAVAVASLIDAHAGLGVIGADVGRHGLICEQOPTYSLRAVLCGTSSCHMGISKDPFV 309
Oy      297 PGVWGPYYSAMVPGLMLVEGGSAGAAIDQLDFHPAVEAREMAQRVNPPLVWLADR 356
Db      310 PGVWGPYFSAMVPGFMLNEGGSQVTKLIDHMVQGHAAPELDQKATARQSIYAYLNSH 369
Oy      357 I-LEKTAQPSDAVALAKGLHAVPEFLGNRPADPRPARAVYICGLGMRDIDLNLAYIAG 415
Db      370 LDLIKRAQF--VGFLTVDLHWMPDFHGNRSPLADLTCLKGWTGLKSQDLDDLAIILYAT 427
Oy      416 LCGIGYGLROIIDAQTAQGVVSKNIYISGAGOHPLVROLADPTGCPVITTCOCEPVLL 475
Db      428 VOAILGSTRITTEAMEAGHSISTLFLCGLSKNPLVONHADITGMPVVLVSQEVESVLV 487
Oy      476 GSAILGAVAGNIAPSVGEAMQOFTHYDKYYUPOERYQSLHNRREYAV--KOLQHTAKLL 532
Db      488 GAAVLGACASGDPRASVGEAMAKMSKVYFPRLDQKKYUQVFLKLVENQKETL 545

```

RESULT 4
 ABB70032

ID ABB70032 standard; Protein: 548 AA.
 XX
 AC ABB70032:
 XX
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 36888.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW.
 XX
 DR WPI, 2001-656860/75.
 DR N-PSDB; ABL14135.
 XX
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 36888; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX
 SQ Sequence 548 AA:
 Query Match 35.8%; Score 1005.5; DB 22; Length 548;
 Best Local Similarity 39.4%; Pred. No. 1.5e-90;
 Matches 215; Conservative 83; Mismatches 216; Indels 31; Gaps 6;
 QY 8 IGVGVGSGSVRAGIFDLNGLSLSHATEKITTTTRSGSRVSSOEIWOAVCSIRNALTL 67
 DB 8 VGVGVGSGSARAALVACDGRLEQAVQITQWNPPEGYINSSNINQSIQVKK--VI 65
 QY 68 ADVCAQSVAGIGDPATCSLVVLDKNGDPLVSPSPGAKONITVMDHRAEQAEIRINATH 127
 DB 66 GGVYDKSKVYKIGIFDPCATSLVVLPGQSPPLTVYSKSGEAEONITIMDRAREDETQEINAFK 125
 QY 128 HPVLNYYGSGISPEMPEPKILMLKENMPEYERAGQFFDLADLTMTATGGLASVCTVT 187
 DB 126 HSLKTYGGGVSLMEVYPKLMLKRNLSQTFGNIRVFDLPDLTWRATGVDTSLCSV 185
 QY 188 CKWTLAHEMRKMPDPYFRITGLAELEDEFRIGHIIVSPCTPGCNLTQAQAAEMGLLP 247
 DB 186 CKMWYDAANSWMKEFLKQADLELTQNNFEKLSDVOPRGRTGKCLITKAAAGELLSA 245
 QY 248 GTPVANGLIDAHAGIGTIVG-----VEGALINLNAVYFGTSSCTMASTTSPSFV 296
 DB 246 GTVAVSTSLIDAHAGALCMFCGRSKESKADVDVG---KNALLAGTSTCHMSITRKACFA 301

QY 297 PGVWGPYYSAMVPGMLWVGSGSAGAAIDQLDFHPAVEARE-----MAQRVNOPL 349
 DB 302 QGVWGPYQDAIIPGFVLENGSGSIAGHLLDHVLSKSHESYAEKLGEDKRFYQHLNKL 361
 QY 350 PVLWADRILEKTAQPSDAVALAKGLHYVPEFLGNRAPADPHAAVYICGLMENDLNL 409
 DB 362 PELAAARGL-----SOVGCITQDVHWPDLGNRSPIADPTLGLGVITGLDMTGTSLA 415
 QY 410 ALYTAGLCIGIGYGLROIADQTAQG-VYSKNIVISGAGQHPVLVQIADPTCGIPVITQ 468
 DB 416 IKYLAFYQALAYGTRHIENLYOYGRAPFQTLCCGGLAKRPVYQCAQADCNLPALIPD 475
 QY 469 CCEPVLGSAITLGAVNAGNIAPSYGEAMQGFTHVDKYYYPQERYOSLHHRRYEAQLOHT 528
 DB 476 EOEWVLVGAALGAAGASGHFDSLSSASKSMGTGOLVKNPNAETLEFHNKKYKVFLOLLEN 535
 QY 529 AKLLR 533
 DB 536 QROYR 540
 RESULT 5
 AAB93914
 ID AAB93914 standard; Protein: 246 AA.
 AC
 AC AAB93914:
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:13884.
 XX
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300283.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oka T, Isogai T, Nishikawa T, Hayashi K, Satou K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 13884; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 246 AA:

Query Match 23.9%; Score 672; DB 22; Length 246;
Best Local Similarity 52.0%; Pred. No. 5,5e-58;
Matches 127; Conservative 38; Mismatches 67; Indels 12; Gaps 4;

OY 110 VMMDHRAEQERINATNHPVLYVGGKISPEMPTKILKEMPEI-YERAGQFEDLA 168
Db 1 MWLDHRAVSQVRINETKHSVLQYGVGVSVMQAPKLLMLENLEICMDKAGHFFDLP 60
OY 169 DELTWRATGDLARSVCTVCTKWTMLAHENRNDPDRFTIGLAELEDEDFIRIGHIYVPG 228
Db 61 DFLSKMAGCVTARSLCSLVCKWTYSA-EKGWDSFEWKMIGLEDEYADNYSKIGNOVLPG 119
OY 229 TPCGNGLPAAQAAEMGLPCTPVAVGLIDAAGIGTGV--VEGAL-----NNLAY 278
Db 120 ASLGNGLTPEARLDGLPGIYAVASLIDAHAGGLGIVADRGHGLICEGPRVTSRLAV 179
OY 279 VEGTSSCTMASTSPSPVPGVWGPYYSANVPGMLWVEGQSAAGAIDQLDFHPAVEEA 338
Db 180 ICGTSCHHGISKDPIFVPGVWGPYYSANVPGFMLEGQSVTGKLIIDHVOGHAFPEL 239
OY 339 REMA 342
Db 240 QVKA 243

RESULT 6

AAU39145
ID AAU39145 standard; Protein: 554 AA.

AC AAU39145;

DT 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #41.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

PA Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
DR WPI: 2001-616774/71.
XX N-PSDB: AAS59506.

PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -

XX Example 1; SEQ ID No 340; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 554 AA:

Query Match 19.7%; Score 553.5; DB 22; Length 554;
Best Local Similarity 28.7%; Pred. No. 1.1e-45;
Matches 154; Conservative 85; Mismatches 227; Indels 71; Gaps 12;

OY 7 VIGVDVSGSVRAGIFDLNGLSLSHATEKITTTTRRSGSRVSGSOETMCAVCSINALLT 66
Db 28 LIGIDVGTSCRAVAIPDLGRPLTFATPRYKTTNPPRGMAEDPDEEMWKNALQASCHRAIA 87

OY 67 LADVCAOSVAGIGFDA-TCSLVVLDKNGDPLPVSPEGDAKONTIWMMDHRAEQERINA 125
Db 88 AAGISPAALINGISYDAITTLTVAMDERGNL-----RPMAMMDVATQAAARAEN 138

OY 126 THHPVLYVGGKISP--EMETPKILKEMPEIYERAGQFEDLDFLTWRATGDLARS 182
Db 139 SDSVARLYNGAGVSPATAEYRPFKAAMLRDHEPETYRRAAHLVDADWDVTFKLTGEWTTN 198

OY 183 VCTVCTKWTMLAHENRNDPDRFTIGLAELEDEDFIRIGHIYSPGCGNGLTAAQAAE 242
Db 199 INSAAIRMYNNRDGGMPEDEFEYETIG---CDVFPKIPGRVLDLGTIPVGT-LGTTPAQL 253

OY 243 MGLLPCTPVAVGLIDAAGIGTVGEGALNNLAYVFGTSSCTMASTSPSPVPGWCP 302
Db 254 LGRPGIPIVAQGLGDMWAGIG-LGVL--APGSMALLTGSNHYLTGOSDTEIHGEFGGA 310

OY 303 YYSANVPGMLWVEGQSAAGAIDQLDFHPAVEEARMAORVNOPLPWLADRIILEKTA 362
Db 311 YSDAVNMPGOYTVGEGSQVSTGSLK-----WFKNFPADYT 345

OY 363 QPSDAVAL-----AKGLAVPEFLGNRAFPDHPARAVYCGIGMERDLD 406
Db 346 AAEKIGLNPYDVLYNQSRIIRPSDGLIINEFQGNRRPYTTSKARGLITGLSLNHTPA 405

OY 407 NLALYIAGICGIGYGRQLDQATOGVYSKNIIVISGAGQHPVLRQLADCGIPIVT 466
Db 406 HFYHAIQESVC---YGTANLRAMKKAAGFEVDKAVACGATKSRDNIOMADVPGPITV 462

OY 467 TQCEPVLGSAALGAVNAGNIAPSVGEAMQOFTHVDKYYP-OERYQSLHRRREAY 522
Db 463 TEVGDAVVLGCTCWAAVAGGLFKDLPEAATQVHEIDFIPDQER-----HEEYQY 514

RESULT 7

AAW53950
ID AAW53950 standard; Protein: 560 AA.

XX AAW53950;
 AC 17-AUG-1998 (first entry)
 DT
 XX
 DE Bacillus subtilis L-ribulokinase.
 XX
 KM Promoter: ara operon; heterologous gene expression; L-arabinose;
 KM arab; L-ribulokinase.
 XX
 OS Bacillus subtilis.
 XX
 PN W09811231-A1.
 PD 19-MAR-1998.
 XX
 PF 10-SEP-1997; 97MO-US16097.
 XX
 PR 10-SEP-1996; 96US-0031077.
 XX
 PA (UYRQ) UNITV ROCKEFELLER.
 PI De Lencastre H, De Saenquelra I;
 XX
 PI WPI: 1998-207392/18.
 DR N-PSDB; AAV25568.
 XX
 PT L-arabinose-utilising operon promoter - useful for regulation of
 PT heterologous protein expression in prokaryotic host cell
 XX
 PS Example 1; Fig 2A-B; 83pp: English.
 XX
 CC This polypeptide comprises the L-ribulokinase of Bacillus
 CC subtilis, encoded by the arab gene of the ara operon (see AAV25568).
 CC This enzyme is involved in the utilisation of L-arabinose by
 CC B. subtilis. The invention relates to the strong promoter (see
 CC AAV25567) of the ara operon, which is induced by L-arabinose and
 CC repressed by glucose, and its use in regulating the expression of
 CC heterologous genes in prokaryotic hosts.
 XX
 SQ Sequence 560 AA;
 Query Match 16.0%; Score 448.5; DB 19; Length 560;
 Best Local Similarity 27.1%; Pred. No. 3e-35;
 Matches 133; Conservative 95; Mismatches 237; Indels 79; Gaps 20;
 QY 8 IGVNDSGSGVRAIGFDL-NGSLSHATEKI-----TTTRSGSRVE-----OSSQEIWQ 55
 DB 5 IGVDFGLSGRAVLVHQTGEELAAVKEYRHAVIDTVLPRTGQKLRDMALQHPADYLE 64
 QY 56 AVSCIFRNATLADVCASQAVAGTGFNT-CLVYVLDKNGDPLVYSPRGDAKON--ITVMM 112
 DB 65 VLETTTPSLLEQGTGVDPKDIDIGIDTACTTIPIDSSGQPLCLPVEEPHPSYVWLK 124
 QY 113 DHRATEQAEKIN-----ATHHPLVNVGKTSPEMERPKIIMKENMEIYERAGQFFDLA 168
 DB 125 HHAQAOKADLNLQAEEGGAFIQRIGGKISSEMMIRKQMLEEAHITIEADRIIEAA 184
 QY 169 DFLTWRAIGDILARSVCVTCTKMTWLAHENRMDPDYFTIGLA-ELADEDFIRIGHIYSP 227
 DB 185 DWLVYQDGLSKRSNCTAGKAKMSEKAGPSDDFEFKLMPSKTTTKD--KISGSHSY 242
 QY 228 GTPCGNGLTAQAAAEMGLPRTPAVAGLIDAHAGIGTGV-EGGALNINLAUYFGTSSCT 286
 DB 243 GKEAGS-LFEKMAKGLGLPSTAVAVANVAHV-SVPAVGIETEPG---KMLMNGSTTCH 297
 QY 287 MASTTSPSPYPCWGPPIYSAMVGLMLVEGGGSAAGADQLLD--FHRAV-EEARE--- 340
 DB 298 VLLGEEHYHIVPGMGVVDNGLTPGAYGEAGGSCVGHDFMFWKTCVPAYOEAKENI 357
 QY 341 -----MAORVNOPPLVWMLDRILEKTAQPSDVAALAKGLHVPEFLGNRAFPDPRHRAV 395
 DB 358 GVHELLSEKANHQAFC-----ESGLLALDMMWNGNRSTLVADLTGM 398

QY 396 ICGIGMERDLDNLALYIAGLCGIGYGLRQILDQAOTAGVSKNIVISGG-AGOHPLVRQ 454
 DB 399 LIGMTL---LTKPREIYRALVEATVGRMIIEFKESGVIEELFAAGIAENKPFVWQ 455
 QY 455 ILADTCGIPVITTOCCERVLVLSAIIAGVAG-----NIAPSVGEAMQOFTIYDK 503
 DB 456 IYADVNTMMDIKISGSPAPALGSAIFGALAGKEKGGYDIDIKAAAMNGKLK-----DI 509
 QY 504 YVYPOERYOSLHRRYEAAYKOLQ 527
 DB 510 TYTPNENAAVYKELVAYKELVH 533
 RESULT 8
 AAU39179
 ID AAU39179 standard; Protein: 593 AA.
 XX
 AC AAU39179;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #75.
 XX
 KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KM dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 PN W0200181581-A2.
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001MO-US12665.
 XX
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skelky YW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'atsosonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI: 2001-616774/71.
 DB N-PSDB; AAS59506.
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 PS Example 1; SEQ ID NO 374; 1069pp: English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 593 AA:

Query Match 15.5%; Score 435.5; DB 22; Length 593;
 Best Local Similarity 26.9%; Pred. No. 6,5e-34;
 Matches 154; Conservative 79; Mismatches 263; Indels 77; Gaps 17;

```

QY 7 VIGVDVSGSVRAGIFDNLNSL-----LSATEKITTRSGSRVE-----QSSQET 53
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 30 VVGLDGTLSGRAVVRANDGAEMGAAYHEYPGVMDRTLSAADGRKLPPDFALDPVDY 89
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 54 WQAVCSIRNALTADVCAQSVAGIGFDPATCSLVV-LDKNGDPLPVSPE--GDAKQNIIV 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 90 LETLETIVGAVVDAGVDPRPHIVIGLDVTSATVVAATKDGTCVLCOLPERRNPRAVVKL 149
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 111 WMDHRAEQEAERI---NATHRPVLYVGKISPEMETPKIIMLKEMPEIYERAGQFPD 166
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 150 WKHHGAQDQADRIYKLAQVREPEMLTRYGILSSEMIMPELTLEBAQVYRATQVFCN 209
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 167 LADFLWRATGDLARVSVCTCKMTLHNNRM-DRDYFTTIGLAEADDFRIGHIY 225
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 210 VLDMLTWRLLTGLVLAFS--TGDSGYKRYQDGKYSRDYLMNINL-PEFADY---FAEKM 262
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 226 SPGTGCG---NGLTQAAMAEMLPGTPVAVGLIDAHAGIGTVGEGALNNLAYVFGT 282
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 263 APVPLPGARVAGLTPEFSEKTLGPVGTVAAGNIDAHVTAANQAVENG---QMTALMGT 319
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 283 SSGTMASTTSPSPFVPGVWGPYSAWVPGMLVVBGGOSAGACAIDQLD-----FHPAV 335
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 320 SACYVVPQGLKKEVPGMFGVGDGIVDGSWGFEAGTAVGDIFAMFIDNCVPGSYFEAD 379
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 336 EEARMAQRNOLPWLMDRIIEKTAQPSDAVALAKGLHVPFLGNRPAPDPRARAV 395
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 380 HRC-----IGVVDLLEKCARQEVG---AHGLIALDMHNGNRSVLADANLSGM 424
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 396 ICGIGMERDNLALYIAGLCIGYGLRQILDQAQGVASKNIYVSGAGOHPLVRQI 455
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 425 ILGQTLTTPED---QYRALLESTAFGAKRTIISFRDSGVEINELVAVAGLTKNTFLMQL 481
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 456 LADTGSIPVITTTQCCPEVLLGSAIILGAVAGNIAPSVGEAN-----QOETH 500
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 482 FCDICHRPLSVGTIKQPGAGSAYFAAVALDLYPVAASAGAKKAGYQYIDEQRAEQ 541
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 501 VDKYVYPRQERYOSLHHRVFAVYKQLOHTAKLLR 533
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 542 YDALYV---EYARLHDYFGRGNQVMHRLKEIR 571
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 9
 AAB93394
 ID AAB93394 standard; Protein; 252 AA.

```

XX AAB93394;
XX AC
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:12575.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.

```

PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR MPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs
 XX
 PS Claim 8: SEQ ID 12575; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification; The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 252 AA:

Query Match 15.4%; Score 434; DB 22; Length 252;
 Best Local Similarity 39.5%; Pred. No. 2.4e-34;
 Matches 98; Conservative 45; Mismatches 101; Indels 4; Gaps 3;

```

QY 287 MASTTSPSPFVPGVWGPYSAWVPGMLVVBGGOSAGACAIDQLDPRNAYVEAREMARVN 346
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 1 MGISKDPEIFVPGVWGPYSAWVPGFMLEGOSVTKGLIDHMGVGAAPFELQVKATARC 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 347 QPLPVMILADRI-LEKTPQOPSDAVALAKGLHVPFLGNRPAPDPRNRAVYICGEMERDL 405
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 61 QSTIYAVYINSHLDIKKAP--VGFITYVDLHWDFPHNRKRPPLADLTGLKAVNTGLKLSQDL 118
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 406 DNLLALYIAGLCIGYGLRQILDQAQGVASKNIYVSGAGOHPLVRQIADTTCGIPVT 465
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 119 DDLALYIATVQAIALGTRITLIEAMEAGHSITLFLCGLSKNRPLRYQNHADITGMRYV 178
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 466 TTQCCPEVLLGSAIILGAVAGNIAPSVGEAMQOFTNHDYKTYRPOERYOSLHNRKYEAL-KQ 524
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 179 LSGEVEVILGAVAVLGAACAGSDFAVSQDEAMAKSKYGVVPRLODKKYKQYVELKL 238
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 525 LOHTAKLL 532
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 239 VEHOKETL 246
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 10
 AAB939224
 ID AAB939224 standard; Protein; 518 AA.
 XX
 XX AAB939224;
 XX AC

DT 25-JUL-2002 (first entry)
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4069.
DE Staphylococcus epidermidis: open reading frame: ORF; bacterial infection;
XX Staphylococcus epidermidis: gene therapy.
KM antibacterial.
XX Staphylococcus epidermidis.
OS US6380370-B1.
XX US6380370-B1.
XX 30-APR-2002.
XX 13-AUG-1998; 98US-0134001.
XX 14-AUG-1997; 97US-055779P.
XX 08-NOV-1997; 97US-064964P.
XX (GENO-) GENOME THERAPEUTICS CORP.
PA Doucette-Stamm LA, Bush D;
XX WPI: 2002-381255/41.
XX N-PSDB: ABN91769.
DR Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
PS Disclosure; SEQ ID 4069; 267pp; English.
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP3124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
SQ Sequence 518 AA;
Query Match 12.1%; Score 340; DB 23; Length 518;
Best Local Similarity 25.4%; Pred. No. 1.6e-24;
Matches 137; Conservative 85; Mismatches 236; Indels 82; Gaps 19;
QY 25 NGSLSHATEKITTRRGSRVDSOIRINQAVCSIRNALTLDYCAQSVAGIGFDATC 84
DB 18 NGTWNLSLYDK----PLPGNTFLQADYDQLEQGVQFLEDSKVNKNNDVVGIGVDT 73
QY 85 SLVY-LDKNGDPLVSPGDAKON---IIVMDHRAEQAEKRI--NATHHPLVLYVCGK 137
DB 74 STIFLDQFRL--HREDKATPNHAYVKMKHHGADENAYIMQSKNNKMLDYGSS 131
QY 138 ISPEMERPKIIMLKENMEIYERAGQFADLTWRAATGDLANSVCTVTKWTWLAHEN 197
DB 132 VNSEMMIPKILEVKNHEAPEILRRARYIEMAGDYITSLTNSINSCIGKGFW-DNEA 190
QY 198 RMDDYPTITLADLDEDFIRIGHN-----IVSPGTCCGGLTRQAAMENGLRGTRYA 252
DB 191 GFNDYDFHSV-----DPLDKIVKEKCEAPDISGESAGR-LCKDYQOILSDQDVVS 243
QY 253 VGLIDAHAGIGTGVVEGALNLAUVFGTSSCTMASTTSFVGVWGPYYSAMVPLTW 312
DB 244 PFIIIDAHSGVGVGAIENG---EFTAVIGTICHLMDSRVYISITGSVKNAIIPFLY 300
QY 313 LVESGQSAAG-----AAIDQLDFH-PAVEARMAQRVNOPLFVWLADRILE 359
DB 301 AVEAGQPAVGDLEFYSKNQAPKRIHQANENHMPVLNLELASHI-----RIEE 350
QY 360 KTAQPSDAVALAKGLHVPERL-GNRAPFADPHARAVYCGIGMERDNLALYIAGLCG 418

DB 351 QHV-----VYLDWLNGRSILSNHLSIFGLLQPYEMIRAYTE---A 394
QY 419 IGYLRQILDAQTACGVYSKNIVISGAGQ--HPLVROLADCCIPVTTQCCCEVLGGS 477
DB 335 TAFGTRKIMKQFEDNHIFVHYHYVASGIPQSKLVEIYANLKNRRVYIDSSNASALGA 454
QY 478 AILGAVAGNIAPSVGEA---MQQFTNVDKYYYPOERYSLHRRYEAAYKQLOHTAKLLRD 534
DB 455 AMLGAVGNNAVSTLKEALSMKQ-----PIAYIGEPLEIQKVQAVKPLYHKYCELHD 505
RESULT 11
AAG82039
ID AAG82039 standard; Protein; 521 AA.
XX AAG82039;
AC AAG82039;
XX 03-SEP-2001 (first entry)
DT 5. epidermidis open reading frame protein sequence SEQ ID NO:1172.
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis.
XX Staphylococcus epidermidis.
XX WO200134809-A2.
XX 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US30782.
XX 09-NOV-1999; 99US-0164258.
XX (GLAX) GLAXO GROUP LTD.
XX kimerly WJ;
XX WPI: 2001-316495/33.
XX N-PSDB: AAH52889.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 18; Page 337; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce host cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the amplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
SQ Sequence 521 AA;
Query Match 11.8%; Score 330.5; DB 22; Length 521;
Best Local Similarity 25.1%; Pred. No. 1.4e-23;
Matches 135; Conservative 86; Mismatches 252; Indels 65; Gaps 17;

```

OY 25 NSLSLHATEKITTTTTRRSRVE-----OSSOEIMQAVCSICRNALLTLADCAQSV 75
DB 8 NGHIISRYEDYANGTYMNSLYDKPLPENYFLIONADYQILEOGVQFVLEDSKKNND 67
OY 76 AGIGDATCSLVV-LDKNGDPLPVSPEDAKON----IIVMDHRAEQAEIRI--NATHH 128
DB 68 VGIGVDFSTSTIIFLEDEPEPL--HRHEDLKTNPNAVYVLMKHHGADENYVIGSKNK 125
OY 129 PVLNVYGGKISPEMETPKILMLKENPEIYERAGOFFDLADFLTWATGDLARSVCTVYC 188
DB 126 NMLDYVSSVNSEMMIPIKILEVKNHAPILRRARYIMEAGDYITSTILTSNINRSNGIGF 185
OY 189 KNTWTLNHRMWDYFRTTIGLAELADEDFIRIGHN-----IYVPTPCNGSLTAQAAM 243
DB 186 KGWV-DNEGAFNDFPHSV-----DPDLPKYKECEAPITISIGSAGR-LCKDYQOIV 237
OY 244 GLLPGTPVAVGLIDAHAGGIGTVGVGALNMLNAVYEGTSSTMASTTSPSFVPGWGPY 303
DB 238 GLSQVYQVSPFIIDAHSGVGLVGAIKAG--EFTAVIGTSTCHMLDLSRQVPISITGSV 294
OY 304 YSAMVNGILVLEGGSAAAGAIQDLDFHR--AVEEAREMAOVNOLPVLADRILEKT 361
DB 295 KNAIIPGLAYEAGOPAVGDLFEYSKNQAPKHIIVDQANEHMHV-----LNLY 342
OY 362 AOPSDAVALAKGLHVPPEFL-GNRAPFADPHARAVICGLMERDLNLLALYTAGLCIG 420
DB 343 ELASIRIREOHVYVLDMLNGNRSLNSHLTGSIFGLTQPRYEMIHAYIE---ATA 399
OY 421 YGLROIILDAOTAGVSVKNIVISGAGQ--HPLVROIILADTCGIPVITTOCCEPVLGSAI 479
DB 400 FGKRLIMKQFEDNHRIVHFFVYASGAIQKSKILLVEIYANVLRKRVYIDSSNASALGAM 459
OY 480 LGAVNAINAPSVGEA---MQQFTNVDKYYTPQRYOSLHNRKREAKQOLHTKLLRD 534
DB 460 LGANVNAVSTKEALSMKQ-----PIAYIOEPLOKQVAKPKYHKYCELLD 508

RESULT 12
ID AAM42079 standard; Protein; 164 AA.
AC AAM42079;
XX
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 7010.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX MO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA

```

```

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dymnac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI61235.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 7010; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 164 AA:
XX
XX Query Match 11.7%; Score 330; DB 22; Length 164;
XX Best local Similarity 50.4%; Pred. No. 2.6e-24;
XX Matches 65; Conservative 20; Mismatches 42; Indels 2; Gaps 1;

OY 8 IGVVSGSVYRAGIFPLNGSLSHATEKITTTTTRSGSRVSOSSOEIMQAVCSICRNALLTL 67
DB 38 VGVDVGTGSVRYALALVDQSGVLLAFADQPIKNMPEQFHHHDSSEDDIACCVYTKKYVG 97
OY 68 ADVCAQSVAGIGFDTATCSLVVLDKNGDPLPVSPEDAKONIIYVMDHRAEQAEIRINATH 127
DB 98 IDL--NQIRGLGPDATCSLVVLDKOFHPLPVNOGDSHRNVIMWLDHRAVSQVNRINETK 155
OY 128 HPVLNVYVG 136
DB 156 HSYLQYVGG 164

RESULT 13
ID ABP30486 standard; Protein; 505 AA.
AC ABP30486;
XX
XX 02-JUL-2002 (first entry)
DE Streptococcus polypeptide SEQ ID NO 10148.
XX
XX Streptococcus agalactiae.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX

```



```

OY 8 IGVVSGSVRAGIFDLNGLSLSHAT-EKITTTRSGSRVSOSSOETIMOVASCIRNALP 66
DB 18 LSLIDYGGTITKALIFPKLHQIAVSSFEIKNTQSGHR-QVNLVTWNAITTAIEVIO 76
OY 67 LADVCAQSVAGIGFDATC-----SLVYLDKNBDPLPVSPEDAKONIIYMDHRAEQAE 121
DB 77 ISKLSPEQISAV-----ACIGHGKGLYLDNKLLEPL-----EGGILSTDRANDLAQ 123
OY 122 RINATHHPLVNYGGKISPEMETPKIL-WLKEMPELEYERAGOPFLDPLTRARAGDLA 180
DB 124 YVESKLDNITWELTROHIFPS-QSPVILRLMLKDYPPEYKISGAVLSAKPIRKLTKGYO 182
OY 181 RSVCTVCKMTWLAHENRW-----DP---DYF---RTI---GLAEADE-DFIRIGHNI 224
DB 183 QEVGD-----ASGNHMINFOTGYIDPAIDDFGIRELNSLPLEIDBADLV----- 228
OY 225 VSPGTPCGNLTAQAAAEMLDPTPVAVGLIDAHAGIGTGVEGGALNNLAVFGTSS 284
DB 229 --PG-----GISSQAAKETGLVEGTPVVGGLFDIDACALSGVLESD-----TFEVIS 274
OY 285 CTMASTTSPFVP-----GWMGPYYSAMVPGMLVVEGGSAAAGALDOLD--FHPAVEE 337
DB 275 GTWNIITTPSLKPAKODSGIMTSYFP---DRVILLESASPTSGAGNLNFMKLMHDEIDN 331
OY 338 AREMAORVNOPLPWLADRILEKTAQPSDAVALAKGLHVPEFLGNRAPDPHARAVIC 397
DB 332 AKSSGSIYDNLEPL-----THTDATHHGLIFPEPLXGSN---TSODASACFF 377
OY 398 GIGMERDLDNL-ALYIAGIGIGYGLRQ-ILDAQTAGVYSKNIVISGAGOHPLVRQI 455
DB 378 GLTTKSTKSQIMRAYE---GIAFAHKHITDLIKSRGSVPKIRPSGGATNSPAMQM 433
OY 456 LAPTCCGIVTTCCCEVVLGSAIILGAVAGNIAPSVGEAMQOPTHVDKYTPQERTSLH 515
DB 434 FSOILNFPLETTVEGTGLGGAILARHALD-KISLEAVODMVRAKAIYKPOLSEVKG 492
OY 516 HRRREYAKOQHT 528
DB 493 KKKYHAYOKLLET 505

RESULT 15
ID ABB47468
AC ABB47468;
DB 05-FEB-2002 (first entry)
DE Listeria monocytogenes protein #172.
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
  vitamin B12; bacterial infection; disease.
OS Listeria monocytogenes.
PN MO200177335-A2.
PD 18-OCT-2001.
PF 11-APR-2001; 2001MO-FR01118.
PR 11-APR-2000; 2000FR-0004629.
PA (INSP ) INST PASTEUR.
PI Buchrieser C, Frangoul L, Couve E, Rusnlok C, Psihl H, Dehoux P,
  Dussargot O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P,
  Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
  Dominguez-Bernal G, Garrido-Garcia P, Tlerez-Martinez A, Amend A,
  Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
  Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

```

```

PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
DR WPI: 2002-010914/01;
XX
XX genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides;
XX
PS Claim 6; SEQ ID No 173; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see AB03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 504 AA:
Query Match 11.3%; Score 317; DB 23; Length 504;
Best Local Similarity 23.7%; Pred. No. 2.9e-22;
Matches 132; Conservative 98; Mismatches 240; Indels 88; Gaps 22;
OY 1 MTTTKYIGVGVSGSVRAGIFDLNGLSL-SHATEKITTRSGSRVSOSSOETIMOVACS 59
DB 1 LTKSPYIMGVGDIGSSRKAVLFFNORGVEIVFRQATHYELINDRG-KAESPTLEIDVLT 59
OY 60 CIRNALLADVCAQSVAGIGF-DATCSLVYLDKNBDPLPVSPEDAKONIIYMDHRAE 118
DB 60 SIQ-AVAKNINTEELAGISFSSAMSLIMVSGELL-----TECITADNGSSE 108
OY 119 OAEIRINATHHPLVNY--VGKISPEMETPKILWLKEMPELEYERAGOPFLDPLTRAT 176
DB 109 TLEKVKRDNITLFQLYEATGPIHPMSPPAKICWLKEKPTLKFTEKFVDIKSYILRL 168
OY 177 G--DLARSVCTVCKMTWLAHENRWDPDYFTTGLAELADEDFIRIGHNIVSPGTCGNG 234
DB 169 GVMWMDSLASGTGLYINIMHD--WE--FEAMEIVLT-PDFL---PRVVPETYLAG 218
OY 235 LTAQAAEMLDPTPVAVGLIDAHAGIGTGVEGGALNNLAVFGTSSCTMAST---- 290
DB 219 VKREYALMGIPEMLPFIIGSD---GALANIGIQATGQNDVITVTSAGAVKRLTDQF 275
OY 291 ---TSPFVPGWGPY--SAMYPGMLVVEGGSAAAGALD-QLDFHRAVEARMAQ 343
DB 276 IDSRGRTFCYGAADGYFIAGAVNNGKVVEMGIDQFSAEISRRDFASFTAIIEV-- 333
OY 344 RVNQPPLPWLADRILEKTAQPSDAVALAKGLHVPEFLGNRAPDPHARAVICGLMER 403
DB 334 -----PPG-----AAGLLPQYILGERAPFWTNDIGGFVGL----- 365
OY 404 LDNLALYI-AGLCGIGYGLRQILDAQTAGVYSKNIVISGAGOHPLVRQIADTCGT 462
DB 366 TINHTKAHFIKALLEGAVAFNLAEYEAVSAPDI---IYTGGISADAMCKLADILNR 422
OY 463 PVITTCCEVVLGSAIILGAVAGNIAPSVGEAMQOPTHVDKYTPQERTSLH-----YQ 512
DB 423 EIRVPHITESSIGALAIKRSGLIKDLN--LKHPLIKAVVHPSENVILKVAELRLIFK 480
OY 513 SLHHRREYAKOQHTAK 530

```

Tue Mar 25 12:01:56 2003

us-09-802-208b-4.rag

Db : : | | |
481 QVTTQLMSSYSQNLNSMOK 498

Search completed: March 13, 2003, 16:53:11
job time : 46.7632 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:37:32 ; Search time 14.5877 Seconds
(without alignments)
1077.057 Million cell updates/sec

Title: US-09-802-208b-4

Perfect score: 2811
Sequence: 1 MTITKTVIGVDGSGSVRAG.....HHRRYKAYKQLOHTAKLLND 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/CTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448.5	16.0	560	US-08-926-842B-13	Sequence 13, Appl
2	393.5	14.0	566	US-08-926-842B-18	Sequence 18, Appl
3	368.5	13.1	569	US-08-926-842B-19	Sequence 19, Appl
4	340	12.1	518	US-09-134-001C-4069	Sequence 4069, Ap
5	293.5	10.4	497	US-09-134-001C-4411	Sequence 4411, Ap
6	290.5	10.3	498	US-09-172-952-26	Sequence 26, Appl
7	265.5	9.4	501	US-09-172-952-15	Sequence 15, Appl
8	259	9.2	514	US-09-134-001C-4347	Sequence 4347, Ap
9	255.5	9.1	504	US-09-134-001C-4170	Sequence 4170, Ap
10	243.5	8.7	485	US-09-172-952-27	Sequence 27, Appl
11	167	5.9	709	US-08-968-563-18	Sequence 18, Appl
12	167	5.9	709	US-08-969-683A-18	Sequence 18, Appl
13	167	5.9	709	US-09-297-928-15	Sequence 15, Appl
14	106.5	3.8	4928	US-09-370-700-5	Sequence 5, Appl
15	106.5	3.8	4928	US-09-370-700-5	Sequence 5, Appl
16	104.5	3.7	527	US-08-677-970-3	Sequence 3, Appl
17	103.5	3.7	731	US-08-911-364-1	Sequence 1, Appl
18	103.5	3.7	733	US-08-664-700-2	Sequence 2, Appl
19	102.5	3.6	527	US-08-907-229-2	Sequence 2, Appl
20	102.5	3.6	527	US-09-574-462-2	Sequence 2, Appl
21	102	3.6	792	US-08-678-039A-40	Sequence 40, Appl
22	101.5	3.6	480	US-09-108-020-6	Sequence 6, Appl
23	99	3.5	881	US-08-960-048-8	Sequence 8, Appl
24	98.5	3.5	1334	5476657-1	Patent No. 5476657
25	97.5	3.5	3739	US-09-320-878-2	Sequence 2, Appl
26	97.5	3.5	3739	US-09-105-537-33	Sequence 33, Appl
27	97.5	3.5	5087	US-09-144-085-1	Sequence 1, Appl

28	97.5	3.5	11877	4	US-09-105-537-6	Sequence 6, Appl
29	96	3.4	2475	4	US-09-413-814-48	Sequence 48, Appl
30	94.5	3.4	2152	4	US-09-036-987A-3	Sequence 3, Appl
31	94.5	3.4	2152	4	US-09-370-700-3	Sequence 3, Appl
32	93	3.3	612	2	US-08-746-283-31	Sequence 31, Appl
33	93	3.3	612	2	US-08-746-257A-29	Sequence 29, Appl
34	93	3.3	613	2	US-08-746-283-1	Sequence 1, Appl
35	93	3.3	613	2	US-08-746-283-3	Sequence 3, Appl
36	93	3.3	613	2	US-08-746-283-5	Sequence 5, Appl
37	93	3.3	613	2	US-08-746-257A-1	Sequence 1, Appl
38	93	3.3	613	4	US-09-347-878-48	Sequence 48, Appl
39	93	3.3	700	4	US-09-408-647A-2	Sequence 2, Appl
40	93	3.3	1060	4	US-08-911-393-2	Sequence 2, Appl
41	92.5	3.3	1403	1	US-07-908-253-3	Sequence 3, Appl
42	92.5	3.3	1403	2	US-08-694-865-17	Sequence 17, Appl
43	92.5	3.3	1403	3	US-08-535-837-3	Sequence 3, Appl
44	92.5	3.3	1403	3	US-09-124-491-17	Sequence 17, Appl
45	92.5	3.3	2647	2	US-08-583-562B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-926-842B-13
Sequence 13, Application US/08926842B
Patent No. 6030807
GENERAL INFORMATION:
APPLICANT: Sa-No. 6030807ueira, Isabel
APPLICANT: de Lencastre, Herminia
TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926,842B
FILING DATE: 10-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-089 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
ORIGINAL SOURCE: protein
ORGANISM: Bacillus subtilis
FEATURE:
OTHER INFORMATION: /product= "arab"

Query Match 16.0%; Score 448.5; DB 3; Length 560;
Best Local Similarity 27.1%; Pred. No. 3.2e-37;
Matches 153; Conservative 95; Mismatches 237; Indels 79; Gaps 20;

STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926,842B
FILING DATE: 10-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-089 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ. ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single.
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: *Salmonella typhimurium*
US-08-926-842B-19

Query Match 13.1%; Score 368.5; DB 3; Length 569;
Best Local Similarity 26.8%; Pred. No. 5.6e-29;
Matches 136; Conservative 79; Mismatches 235; Indels 113; Gaps 27;

8 IGVVSGSVAGIFDLSLSLHAT-EKITTTRSGSRVEQSS-----50
5 IGLDFGSDSYRALAVDC-----ATGDELATSVEMPRMOEGYCCGRNNQFRHHRDXY 57
QY 51 QETWQAVSCIRNALTLADYCAOSVAGIGDATCSL-VLDKKGDPILVSPREGDAKON-- 107
DB 58 MESMEALKAULVLAQLSAAQ--RANVVGIGVDSGTSPAPIDADGNNVLAERPEFENFAM 115
QY 108 IIVMMDBRATQEARINATNH-----PVLNVGSKISPEMETKILMLKEMPEIYRAG 162
DB 116 FVLMKDBTAAVEADELTRLCHEKPKYDYSRIIGITISSEFMWALHTVTRQDSAVQAAY 175
QY 163 QFEDLADFLTRATG-----DLARSYCTVCKWTMLAHENRMDPDYFRTIGL--AELADE 215
DB 176 SWIELDQWVALLSGTTRPRODIRGRSAGHKITLM--HES-WG-----GLPPASFDE 225
QY 216 DFRIGHNIIVSP-----GTPCGNGLTAAQAAEMGILLPTPVAVGLIDAHAGIGTVG 267
DB 226 LDPCINHLRYPLFSETFTADLPVGT-LCAEMAQRLDLPESVVISGAFCHMGAVGA-- 282
QY 268 VEGGALNNLAVYVGTSCSTMASTSPSF-----VPVWVGPIYSAMPVRLMVEGOSAGA 323
DB 283 --GAQPTLVKVIQSTCDILADKOSVGRAYKGLIGOVDSVVPFRTIGLENGQSARFGD 340
QY 324 -----AIDOLDLFRPRAVEAREMAQRVNOPLRVMLADRLIEKTAQPSDAVLAK 372
DB 341 IYAMFSVILSWPLEQALAAQRELPQINASOK--QLLPAL-----IDAMAKNP 386
QY 373 GLHNVPE-----FLGNRPAPDPHARAVYIGLGMERDIDNLALYIAGLCIGIGLROILD 428
DB 387 SLPHLPLVLDLWFGRRTPRPNORLKGIVITDNLNLTADAP--ALFGGLVASTAFARAIOE 443
QY 429 AQTAGGVSKNIYISG--AGONPLVROILADTGISP--VITTOCCSEPVLLGSALLGAVA 484
DB 444 CFTDQGIANNVNMALGSIARKNOYIMQVCCDVLNRLQIYASDGC--ALGAAIFRAVA 500
QY 485 GNAPSVGEAMQOF-THVDKYYYPQ-----ERYOSLHRRRYEAY 522

Db 501 AKVHADIPAAQSGMASAVERTLRHPHOAQRFPOL-YRRYQOW 542

RESULT 4
US-09-134-001C-4069
Sequence 4069, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4069
LENGTH: 518
TYPE: PRT
ORGANISM: *Staphylococcus epidermidis*
FEATURE:
NAME/KEY: UNSURE
LOCATION: (6)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-134-001C-4069

Query Match 12.1%; Score 340; DB 4; Length 518;
Best Local Similarity 25.4%; Pred. No. 4.1e-26;
Matches 137; Conservative 85; Mismatches 236; Indels 82; Gaps 19;

25 NSGLSHATEKITTRRSGRVDSQETWQAVSCIRNALTLADYCAOSVAGIGDATC 84
18 NGTYMNSLYDK--PLRGVYFLQNNADYDQLLEQGVQFVLEDSKVNKNVVGIGVDFTS 73
QY 85 SLVY-LDKNDDPLVSPREGDAKON--IIVMMDBRATQEARI--NATHNPLVNVYSGK 137
DB 74 STIIFLDEQEP--NRHEDLKTTPHAYVYKLMKHNGADENYTIQSKKNWLDYTGSS 131
QY 138 ISPEMETPKILMLKENPEIYERAGQFEDLADFLTRATGDLARSYCTVCKWTMLAHEN 197
DB 132 VNSEMMIPIKILEYKHEARIELRRARYTMEAGDYITSLTNSNRNSGIGFPGW--DNEA 190
QY 198 RMDPVRFTIGLELADEDFIRIGHN-----IVSPGTCGNGLTAAQAAEMGILLPTPVA 252
DB 191 GFNYDFEHSV-----DPDLPKIVKECEAPITISIGESAGR-LCKDYQOITMGLSODVQVS 243
QY 253 VGLIDAHAGIGTVGCGALNNLAVYVGTSCSTMASTSPSPVPGWGPYYSAMPVGLW 312
DB 244 PFIIDAHSGVLGVALBAG--EFTAVIGTSTCHMLDSROVPISSTIGSVKNAIIRGLY 300
QY 313 LVEGGQSAAG-----AIDOLDLFRPRAVEAREMAQRVNOPLRVMLADRLIEKTAQPSDAVLAK 359
DB 301 AYEAGQPAVDDLPYKSNQAPKHIVDQANHHMVLNLYLELASHI-----RIEE 350
QY 360 KTAQPSDAVALAKGLHNVPEFL-GNRPAPDPHARAVYIGLGMERDIDNLALYIAGLCG 418
DB 351 QHV-----VLLMLNNGNRSILNSHITGSIIFGLTLQTPYEMIHRAIE--A 394
QY 419 IGVGLROILDAQPAQGVSKNIYISGAGQ--HPLVROILADTGISPITTOCCSEPVLLGS 477
DB 395 TARGTKLIMQFEDNHNLRPHVTIVASGIPQSKLIVETIVANVLRKRVVLDSSNASSALGA 454
QY 478 AILGAVNGNIAPSVGEA--MOQFTHYDKYYYPQERYOSLHRRRYEAYKOLQHTAKLLRD 534
DB 455 AMGANGNMYSTLKEALSMKO-----PLAYIQEPVIGQVQAYKPLYNHYGCLHD 505
RESULT 5
US-09-134-001C-4411

```
; Sequence 4411, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4411
; LENGTH: 497
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4411

Query Match      10.4%; Score 293.5; DB 4; Length 497;
Best Local Similarity 22.8%; Pred. No. 2.3e-21;
Matches 128; Conservative 83; Mismatches 249; Indels 101; Gaps 18;

QY 3 ITKTVIGDVGGSGVRAGIFDLNGLSLSHATEKITTRRSGRVYEOSSQETMAVCSIR 62
DB 3 VKEVVLIDIDGTSKTIIVDQGNVIESVETLKYOEHPGYSEODPNFEATKKGIK 62
QY 63 NALTLDVCAOSVAGIGFDATC-SLVVLDKNGDPLVPSPEGDAKONIIWMHDHATEQAE 121
DB 63 ELIOSTMSDKIVGIGSFGSMHGLVIYDDNGIRL-----RKAILMDTNSIQCR 113
QY 122 RINATHHPLVNYGKISPEME---TPKILMKENMEIYERAGOFDLAFLTRATGD 178
DB 114 QIEDIYERLNY-----NPLIEGFTLPKMLVQOHEBEIMNRVDVFMPLKDYLRCLTQT 168
QY 179 LARSCVITQKMTWLAHEN-RMPDVFRTIGLAELADEDFIRIGHIYPTCPGNGITA 237
DB 169 IHMEY-SDACSTLLENPDNEMTRDVGDTFNIGDIYPP-----LVKSHSYGN-VTS 218
QY 238 QAAAEMLGLPCTPVAVGLIDAHAGIG-----TVGVEGALNLLAYVFGTSCT 286
DB 219 SLAKELDSSDVANYAGGDAAGCAIGAYIHDKSALCSISTGCVL-NVEYQAVTISDS 277
QY 287 MASTSPSEVPGWGPRYSAMVPGMLVEGQSAAGAILDQLDFHPAVEAREMAORVN 346
DB 278 NLH-----LFNHSVPTTYAMGYTLAAGYSLN----- 304
QY 347 QPLPWLADRILEKTA-----QPSDAVALAKGLHVPPELGNRAPRAPHARAVICGL 399
DB 305 -----WLKOTFEENESEELINLAASSKIG-ANGLLFTPYLAGERTPHGDAQIIGSFIGI 358
QY 400 -GMEERDLNLLALYIAGLCIGYGLRQILDQAOTAGVYVSKNIVISGAGOHPLRQIAD 458
DB 359 SGOHTKADFPARAV-----IEGTTYSLYDSIKIMRAGHEMNSITISGGAKSRFLOQAD 414
QY 459 TCGIPVITTOCCPEVLLGSAILGAVAGNIAPSVGEAMQOFTHVDKYYPQ-----ERY 511
DB 415 IFNVOIKRLHHEBPSMGAILAAYGLGWFKTIESVCEAFIKYDEVPEPNENHDLVEOY 474
QY 512 QSLHHRVEAYKQLOHTAKLL 532
DB 475 YSYVEAIYKQTKOL--TADLL 493

RESULT 6
US-09-172-952-26
; Sequence 26, Application US/09172952
; Patent No. 6368793
; GENERAL INFORMATION:
; APPLICANT: Hoch, James
; APPLICANT: Dattois, Veronique
; TITLE OF INVENTION: METABOLIC SELECTION METHODS
```

```
; FILE REFERENCE: 234/191
; CURRENT APPLICATION NUMBER: US/09/172,952
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 498
; TYPE: PR
; ORGANISM: LysK-Ec
US-09-172-952-26

Query Match      10.3%; Score 290.5; DB 4; Length 498;
Best Local Similarity 24.8%; Pred. No. 4.7e-21;
Matches 136; Conservative 77; Mismatches 243; Indels 93; Gaps 22;

QY 3 ITKTVIGDVGGSGVRAGIFDLNGLSLSHATEKITTRRSGRVYEOSSQETMAVCSIR 62
DB 1 MTQVWLIGDGGSWLKKGLYDREGREGVQVRLPCLALSPQGMAREMDALMOCMAVIR 60
QY 63 NALTLDVCAOSVAGIGFDAT-CSLVVLDKNGDPLVPSPEGDAKONIIWMHDHATE--- 118
DB 61 ALLTHSGVSGBOIVGIGISAQKGLFLDKNDKPL-----GNAILSSDRAMEIVR 111
QY 119 --QAEKINATHHPLVN---YVGKISPEMERPKILMKENMEIYERAGOFDLAFLTW 173
DB 112 RMQEDGIEPKLPLRLOTLMTGHPVS-----LRLWLEKHEPERYAOIGCYMMTHDYLRW 165
QY 174 RATGDLARSVTVCKMTWLAHEN-----RMPDVFRTIGLAELADEDFIRIGH---H 223
DB 166 CLTG-----VKCEESINISESLYNNMSLGEYRDLTDWLGIAE-----INHALRP 210
QY 224 IVSPCTPGNGLTQAAAEMLGLPCTPVAVGLIDAHAGIGTVVEGALNLLAYVFGTS 283
DB 211 VVGSAAEIGCE-ITQATALGLKAGTAVVGGTLPVVSFAL-CAGIEDEFTLNA--VNGTW 266
QY 284 SCTMAST---TSPSEVPGWGPRYSAMVPGMLV-EGQSAAGAILDQLDFHPAVEAR 339
DB 267 AVTSGITRGLRDEGHAHPVIGRYVN---DGEFIYHESPTSSGN-----LEMPTA-QMG 317
QY 340 EMAORVNO---PLPWLADRILEKTAQPSDAVALAKGLHVPPELGNRAPRAPHARAVI 396
DB 318 ISPEINQAVASLPRKAGDGLFELFLYGSNA-----GLEMTSGGYQMA-----IHTRAHL 368
QY 397 CGLMEERDLNLLALYIAGLCIGYGLRQILDQAOTAGVYVSKNIVISGAGOHPLRQI 456
DB 369 -----LQAIYE---GVVEFSHMTHLNRMRERTDVHTLRVYGGRAHSDVMQM 413
QY 457 ADTCGIPVITTOCCPEVLLGSAILGAVAGNIAPSVGEAMQOFTHVDKYYPQOERYOSLHN 516
DB 414 ADVSGLRIELRPOVEETGCFGALARVGTGYHNFSEBQRDLRHRPVRLLPDMTAHOLYQ 473
QY 517 RRYEAYKOL 525
DB 474 KKYRYOHL 482

RESULT 7
US-09-172-952-15
; Sequence 15, Application US/09172952
; Patent No. 6368793
; GENERAL INFORMATION:
; APPLICANT: Hoch, James
; APPLICANT: Dattois, Veronique
; TITLE OF INVENTION: METABOLIC SELECTION METHODS
; FILE REFERENCE: 234/191
; CURRENT APPLICATION NUMBER: US/09/172,952
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 501
; TYPE: PR
; ORGANISM: LysK-Ko
```

us-09-172-952-15

Query Match 9.4%; Score 265.5; DB 4; Length 501;

Best Local Similarity 22.7%; Pred. No. 1.8e-18; Matches 129; Conservative 74; Mismatches 224; Indels 141; Gaps 23;

```

OY 8 IGVNDSGVRAGIFDLNGLSLSHATEKITTTTRSGSRVSSOEIOWAVCSIRNALTL 67
DB 9 LGIDCGGTLKAGLYAEHGHGIVQALRTNSPLPGYERDMRQMLMOCAATITAGLLOQ 68
OY 68 ADVCAOSVAGIGFDT-CSLVLDKNGDPLVSPEDAKONITVMMDHRAE-----QAE 121
DB 69 AGVSGQIKGVISAGOGGFLDKQDRPL-----GNALSSDRRLKIVGRMQRD 119
OY 122 RINATHHPVLN---YVGKISPEMETPKILMKENMPETIERAGGFEDLADFLTWATGD 178
DB 120 RLPERLYPRTROTLLMTGHPAS-----LLRWKENEPQRAQIGVMMGHDIYLRWCLTG- 172
OY 179 LARSVCTVCKWTWLAHEN-----RMPDYEFTIGLAELEDEFLRIGHIIVSPGTPC 231
DB 173 -----AKGGEESNISSESNLYNMAMGOYDPRLEWLGIGETI-DSAL-----PYYVGSATC 221
OY 232 GNGLTAAOAAEMGLPRTPVAVGLIDAHAGIGTGVGE-GALNMLAYVFGTSSCTMAST 290
DB 222 GE-ITAAALITGLAGTTPVGGTLPVYSTAL-CAGIEDESTLNAAVGTWAVTSGIAHGL 279
OY 291 TSPSFYGVWGPYYSAMVGLMVEGGQSAAGAIIDQLDFHRAVEAREMARVNOPLP 350
DB 280 RDHEAHPYVYGRYV--DQYIVHEASPTSSGNLE----- 312
OY 351 VWLADRILEKTAQPSD-----AVA-LAKG---LHVPEFLGNRAFPADPHARAVICG 398
DB 313 -WF-----TAQMDLSPETINQAVASLPKAGSELPFLPLXGNSA-----G 352
OY 399 LGMERDLNMLALYIAGLCIGYGLROI-----LDAQTAQGVSKNIV----- 441
DB 353 LEM-----TCGF-YGMQALHTRAHLLOAVEGVVFSHTHLRMRERTNV 397
OY 442 -----ISSGAGHPHVRQIILADTCGIPVITTOCEPVLGSAIIGAVAGNIAPVGEAMQ 497
DB 398 QALRTVGGRAHSDVMQMLADVSGRLTELPKVEETGCGFALAAVGTGYRSPSEARRA 457
OY 498 FTHVDKYYYPOERYOSLHRRREAYKOL 525
DB 458 RQHPVRTLRLPDMTAHARQYKRYKHNYLHL 485

RESULT 8
US-09-134-001C-4347
; Sequence 4347, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4347
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4347

```

Query Match 9.2%; Score 259; DB 4; Length 514;
 Best Local Similarity 22.3%; Pred. No. 8.5e-18;
 Matches 124; Conservative 82; Mismatches 232; Indels 118; Gaps 21;

```

OY 5 KTVIGVDVSGSVRAGIFDLNGLSLSHATEKITTTTRSGSRVSSOEIOWAV 57
DB 3 KVMIGVDMGTTSKAVLYDENCKFIYKHNIGYDLHPNDVS-----FENDELFDAY 55
OY 58 CSCIRNALTLADVCAOSVAGIGFDT-C- -SLVLDKNGDPLVSPEDAKONITVMMDHRA 116
DB 56 LMTVKYIVRESGIADKIDIFISLAQMHSILIANEDNORLTEN-----ITMDNRA 106
OY 117 TEOAERINATH--HPVLNVYVGKISPEMETPKILMKENMPETIERAGGFEDLADFLTW 174
DB 107 NDVADLIEKSYGFEELYORTGPIHPSPLSKIFEMRHEBKIFKQTA MFADTKTVLLRQ 166
OY 175 -----ATGDLARSVCTVCKWTWLAHENRMDPDYERTIGLAELEDEFLRIG 221
DB 167 LYREFVIDYSLGATGMN-----LEOLDMNAKALELIGIENQPOLY--- 210
OY 222 HHIVSPGTPCGNGLTAAOAAEMGLPRTPVAVGLIDAHAGIGTGVGEALNMLAYVEG 281
DB 211 -----PTTHVLTGMKKRYATLMGIDQTPVIVGASD---GVLSNLGVNSYOKGEVAVTIG 262
OY 282 TSSCTMASTTSPSFYGVWGPYYSAMVGLMVEGGQSAAGAIIDQLDFHRAVEAREM 341
DB 263 TSGAIFRTVINQPK--TDEKGRIFCYLLDKQYVIGGVNGGVVLR----- 306
OY 342 AQRVNQPLPWLADRILE--KTAQ-----PSDAVAL-----AKGLHVPEFLGNRA 385
DB 307 -----WLPDELLASVEVETAKRLGVDPYVLVLIQASRVKPGAEGLFHHYLANGERA 356
OY 386 PRADPHARAVICGLGMERDLNMLALYIAGLCIGYGLROI-LDAQTAQGVSKNIVISG 444
DB 357 PLNMADARGSFELTTLTKKHEHMR--AALEGVLYNLVYVLTALIEVMNEFTTITKATG 413
OY 445 GAGQHPHVRQIILADTCGIPVITTOCEPVLGSAIIGAVAGNIAPVGEAMQFTHVD 502
DB 414 GFAKSETRWQMDIFDTDLIVPESTYSSCLGACVLGAKALGELIDPSTIKDMVGTTHA- 472
OY 503 KYRYPOER---YQSL 514
DB 473 --HEPNEETVAITQOL 486

```

```

RESULT 9
US-09-134-001C-4170
; Sequence 4170, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4170
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4170

```

Query Match 9.1%; Score 255.5; DB 4; Length 504;
 Best Local Similarity 21.0%; Pred. No. 1.9e-17;
 Matches 117; Conservative 87; Mismatches 245; Indels 107; Gaps 20;

```

OY 5 KTVIGVDVSGSVRAGIFDLNGLSLSHATEKITTTTRSGSRVSSOEIOWAVCSIRNA 64
DB 8 KYILSIDQGTSSRAILFNKEGELKVSQREKQHPHPGWVEIDANIEITVSLVWAEL 67
OY 65 LTLADVCAOSVAGIGF-DATCSLVLDKN--GDPLVSPEDAKONITVMMDHRAEQAER 122

```

```

Db 68 LNNNNANQIEGIGITNORETIVVWKNKTRPI-----YHAIYQSRQIDICTN 118
QY 123 I-NATHHPLVNYGKISPEMPTKILMKENMEIYERAGQ-----FFDLADFLTWBRAT 176
Db 119 LKEQYEETFEKREKGLLDLPYFAGTKVKWILDHVGAREKANGDLDLFGTIDSMLVKMTLS 178
QY 177 GDLARSVCVCKMTWL--AHENRMDPYFRTIG-----LAELEDEDI---RICHNIYS 226
Db 179 GRTHAITDYTNASRTLMENIYDLKWDDELLELNI PKOMLEPEVKSSEIYKRTIDYHFEFG 238
QY 227 PGTPCGN-----GLTPQAAEMGLPGTPVAVGLIDAHNGIGITVEGGLNNLNAV 279
Db 239 QEVFIAGIAGQQAALBEGQACFDKGDVKNYGTGSGFMLMNGE--EAVASEGGLTTIY- 296
QY 280 FGTSSTMASTSPSPVPGVWPYYSANVPGMLVVEGQSAAGAIQDLDLPHAVEAR 339
Db 297 -----GLDGKVNVAL-----EGSIFVSGAIDQ----- 318
QY 340 EMAQRVNOPLFVWLAB--RIIEKTAQP--SDAVALAKGLHVPEFLGNRAPFADPHARA 394
Db 319 -----WLRDGLRMINSAPOTENYASRVESTEGVYVWPAFVGLGTPYWDSEARG 366
QY 395 VIGGLMERDLDNLATYIAGIGIGVGLRQILDQAQTA--QGVVSKNIYISGAGOHPLVR 453
Db 367 AIFGLSKTEKEHFIRATLESCL--YOTRDYMEAMSKDSGIEVONLWVDGAAVANNFTM 423
QY 454 QILADTCGIPYITTOCCPEVLLGSAILGAVAGNIAPSVG-----EAMQOFTHVDKYXXYPQ 508
Db 424 QFOADIYVSSVEREIOETTLGAAYLAGLA-----VGFMDKEDIRERKKTLOTEFRP- 476
QY 509 ERYOSLHRRREAVKQ 524
Db 477 EMDADQRHKLYSGWK 492

```

RESULT 10 US-09-172-952-27

```

: Sequence 27, Application US/09172952
: Patent No. 6368793
: GENERAL INFORMATION:
: APPLICANT: Hoch, James
: TITLE OF INVENTION: METABOLIC SELECTION METHODS
: FILE REFERENCE: 234/191
: CURRENT APPLICATION NUMBER: US/09/172.952
: CURRENT FILING DATE: 1998-10-14
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 27
: LENGTH: 485
: TYPE: PRT
: ORGANISM: LYXK-H1
: US-09-172-952-27

```

```

Query Match 8.7%: Score 243.5; DB 4; Length 485;
Best Local Similarity 23.2%: Pred. NO. 3e-16;
Matches 129; Conservative 77; Mismatches 242; Indels 109; Gaps 22;

```

```

QY 8 IGVVGSVYRAGIFDNLGSLSHATEKITTRRSQSVESQSOEIMQAVCSIRNALTL 67
Db 5 LGIDCGGTFIKALIFDNGTIGTARRRIPITSEKPGIAERDMDELMALCAQVIOKTRQ 64
QY 68 ADVCAQSVAGIGFDT--CSLVYLDKNSDPLVSPGDAKONIYVMDHRAE-----QAE 121
Db 65 SSILPQOIKALIGISAQGAFFLDKNKPL-----GRA-----LSSDRAVEIYQCMQKE 115
QY 122 RINTTHPV---LNVYGKISPEMPTKILMKENMEIYERAGOFDLDPLTWBRATGD 178
Db 116 NIIQKFPYITLQTLMMGPVS-----ILRWIKENEPSRYBOJITILMSHYLEFCLTER 169
QY 179 LARSVCVCKMTWLAHENRMDPYFRTIGLAELADEDFIIGHNIYSP--GTPCGML 235
Db 170 LYCEETNISSENFYNNRKGKIDQLAKLFGITTECDK-----LPIITKSNKIAGYV 220

```

```

QY 236 TAQAAEMGLPGTPVAVGLIDAHAGIGITVEGGL-----NLAAYFGTSSCTMA 288
Db 221 TSRAEQSGLEVEIYVVGGLFD-----VSTALCADLKDDOHLNVLGTWSVVG 270
QY 289 ST---TSPSPVPGVWPYYSANVPGMLVVEGQSAAGAIQDLDLPHAVEARMAQRV 345
Db 271 VTHYIDNQTIPFYVYKPYEK-----NKEITIHASPTISAGNLEWVF 311
QY 346 NO-PLPVMIADRI--LEKTAQPSDAVALAKGLHVPEFLGNRAPFADPHARAAYVCGLM 401
Db 312 NOFNILPNY--DDINHEIKLKPASSSVLPAPFLYSNNKLMQAGF-----YGIQS 360
QY 402 ERDDLNL-ALYIAGLCIGVGLRQILDQAQGVVSKNIYISGAGOHPLVRQILDADTC 460
Db 361 HHTQHILLQAIYE-----GVISLMSHLBRMQVRFPNASTVARTVGTGPAKSEVWQMGLADIS 416
QY 461 G-----IPYITTOCCPEVLLGSAILGAVAGNIAPSVGEMQOFTHYD-KYTPQOETYSGLH 515
Db 417 GMRLEIPNIEETGC-----LGAALMAQESAV-----EISQILNIDRKIFLPDKNQYSKY 467
QY 516 HRYEAYKQLOHTAKL 532
Db 468 QKHRYHLYKFIKALKNL 484

```

RESULT 11 US-08-968-563-18

```

: Sequence 18, Application US/08968563
: Patent No. 6013494

```

```

: GENERAL INFORMATION:
: APPLICANT: CHARLES E. NAKAMURA
: APPLICANT: ANTHONY A. GATENBY
: APPLICANT: AMY (KUNAG-HUA) HSU
: APPLICANT: RICHARD D. LA BEAU
: APPLICANT: SHARON L. HAYNIE
: APPLICANT: MARIA DIAZ-TORRES
: APPLICANT: DONALD E. TRIBBUR
: APPLICANT: GREGORY M. WHITED
: APPLICANT: VASANTHA NAGARAJAN
: APPLICANT: MARK S. PAYNE
: APPLICANT: STEPHEN K. PICATAGGIO
: TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
: TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
: STREET: 1007 MARKET STREET
: CITY: WILMINGTON
: STATE: DELAWARE
: COUNTRY: U.S.A.
: ZIP: 19898
: ADDRESSEE: GENENCOR INTERNATIONAL, INC.
: STREET: 4 CAMBRIDGE PLACE
: STREET: 1870 SOUTH WINTON ROAD
: CITY: ROCHESTER
: STATE: NEW YORK
: COUNTRY: U.S.A.
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.50 INCH DISKETTE
: OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
: SOFTWARE: MICROSOFT WORD VERSION 7.0A
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/968.563
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/030,601
: FILING DATE: NOVEMBER 13, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: FLOYD, LINDA AXAMETHY

```



```

: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: CR-9982
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 302-892-8112
: TELEFAX: 302-773-0164
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 709 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: GUT1
: US-08-968-563-18

```

Query Match 5.9%; Score 167; DB 3; Length 709;

Best Local Similarity 21.0%; Pred. No. 4.2e-08;

Matches 116; Conservative 66; Mismatches 182; Indels 188; Gaps 25;

```

OY 46 VEOSSOEIMQAVSCIRNA-LTLADYCAQSV-----GIGFDATCSLVLDKNG 93
DB 153 VECHPOKLLVNVQCLASSLSLQITNSERVANGLPYKVICGIANMRETTILMSRTG 212
OY 94 DPLVPSPEGDAKONIIVMDHRAE-----QAEKINATGHPVLYVGGKISP 140
DB 213 KPI-----VYGIWMNDTRIKIVRDKWNTSVDRQLQKRGKGLPL-----ST 257
OY 141 EMETPKILMKEMP---EYERAGOFEDLADFLTW-----RATGDLARSVCVTG 188
DB 258 YFSCSKIRWFLDNEPLCTKAYEENDLMFTVD--TWLIYQLTOKAFVSQVTNASRTGFM 315
OY 189 KWTWLAHENR---WDPD-----YFTIGLAELADEDFIRIGH-----223
DB 316 NLSTLYKYNLELFEWGIKDLKILHMPETVSSQYRGDCI-----PDMIMKLDSPKTVL 370
OY 224 --IVSPGTPCGNGLTAQAAEMGLLPETPYA-----VGLIDHAGSIGTVGEGALNN 275
DB 371 RDLVKRNLPIQGCIGDQASAMVGLAYKPGAKCTYGTGCFLLYNGTKKLISOHGALTY 430
OY 276 LAVVFTSSCTMASTSPSPVPGVWGPY---YSAMVPGI---WLVGSGSAGAAIDQL 328
DB 431 LAF-----WPHLOEYGGQKPELSKPHFALEGSVAVGAAYVO--467
OY 329 LDFHPAVEAREMAQRVNPPLPYWLAD--RILEKTAQ---SDAVALAKGHLVHPEFLGN 383
DB 468 -----WLRDNLRLIDKSEVGVFIASVYPSGSGVVPVPAESGL 504
OY 384 RAPFADPHARAVICGLMERDLDNLALVYTAGLIGIGYGLRQLIDAQT--AQGVSKN--439
DB 505 FAYWMPDARATIMGSOFTTASHIARAAGEVC---FQARAILIKAMSSDAFEGSKDRD 561
OY 440 -----IYISGAGQHPLVROLADTCGIPVY-----TQCCPEVLLG 476
DB 562 FLEISDVYEKSPLSVLAVDGMRSRNEWQIADILG-PCVKVRSPTAEC-----613
OY 477 SAILGAVAGNIA 488
DB 614 TALGAALIAANMA 625

```

RESULT 12

US-08-969-683A-18

Sequence 18, Application US/08969683A

Patent No. 6136576

GENERAL INFORMATION:

APPLICANT: GENENCOR INTERNATIONAL, INC.

TITLE OF INVENTION: METHOD FOR THE RECOMBINANT

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International, Inc.

STREET: 4 Cambridge Place

```

: STREET: 1870 South Winton road
: CITY: Rochester
: STATE: NY
: COUNTRY: U.S.A
: ZIP: 14618
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FastSeq for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/969,683A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US97/20873
: FILING DATE: 13-NOV-1997
: APPLICATION NUMBER: 60/030,601
: FILING DATE: 13-NOV-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Glaister, Debra
: REGISTRATION NUMBER: 33,888
: REFERENCE/DOCKET NUMBER: GC 369-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-864-7620
: TELEFAX: 650-845-6504
: TELEX:
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 709 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: GUT1
: US-08-969-683A-18

```

Query Match 5.9%; Score 167; DB 4; Length 709;

Best Local Similarity 21.0%; Pred. No. 4.2e-08;

Matches 116; Conservative 66; Mismatches 182; Indels 188; Gaps 25;

```

OY 46 VEOSSOEIMQAVSCIRNA-LTLADYCAQSV-----GIGFDATCSLVLDKNG 93
DB 153 VECHPOKLLVNVQCLASSLSLQITNSERVANGLPYKVICGIANMRETTILMSRTG 212
OY 94 DPLVPSPEGDAKONIIVMDHRAE-----QAEKINATGHPVLYVGGKISP 140
DB 213 KPI-----VYGIWMNDTRIKIVRDKWNTSVDRQLQKRGKGLPL-----ST 257
OY 141 EMETPKILMKEMP---EYERAGOFEDLADFLTW-----RATGDLARSVCVTG 188
DB 258 YFSCSKIRWFLDNEPLCTKAYEENDLMFTVD--TWLIYQLTOKAFVSQVTNASRTGFM 315
OY 189 KWTWLAHENR---WDPD-----YFTIGLAELADEDFIRIGH-----223
DB 316 NLSTLYKYNLELFEWGIKDLKILHMPETVSSQYRGDCI-----PDMIMKLDSPKTVL 370
OY 224 --IVSPGTPCGNGLTAQAAEMGLLPETPYA-----VGLIDHAGSIGTVGEGALNN 275
DB 371 RDLVKRNLPIQGCIGDQASAMVGLAYKPGAKCTYGTGCFLLYNGTKKLISOHGALTY 430
OY 276 LAVVFTSSCTMASTSPSPVPGVWGPY---YSAMVPGI---WLVGSGSAGAAIDQL 328
DB 431 LAF-----WPHLOEYGGQKPELSKPHFALEGSVAVGAAYVO--467
OY 329 LDFHPAVEAREMAQRVNPPLPYWLAD--RILEKTAQ---SDAVALAKGHLVHPEFLGN 383
DB 468 -----WLRDNLRLIDKSEVGVFIASVYPSGSGVVPVPAESGL 504
OY 384 RAPFADPHARAVICGLMERDLDNLALVYTAGLIGIGYGLRQLIDAQT--AQGVSKN--439
DB 505 FAYWMPDARATIMGSOFTTASHIARAAGEVC---FQARAILIKAMSSDAFEGSKDRD 561

```

OY	440	-----IYISGACOHPLVRQILADTCIPVY-----TTQOCBPVLIG	476
DB	562	FLEELSDTYEKSPSLVLAVDGMSRNSNEVMQIDILG-PCVKARSRFTAEC-----	613
OY	477	SAILCAVAGNIA 488	
DB	614	TALGAAMMA 625	
		RESULT 13	
		US-09-297-928-15	
		Sequence 15, Application US/09297928	
		Patent No. 6358716	
		GENERAL INFORMATION:	
		APPLICANT: BULTHUIS, BEN A.	
		CATENBY, ANTHONY A.	
		HAYNIE, SHARON L.	
		HSU, AMY K.	
		LAREAU, RICHARD D.	
		TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF	
		GLYCEROL BY RECOMBINANT	
		ORGANISMS	
		NUMBER OF SEQUENCES: 25	
		CORRESPONDENCE ADDRESS:	
		ADDRESSEE: GENENCOR INTERNATIONAL, INC.	
		STREET: 4 CAMBRIDGE PLACE	
		1870 SOUTH WINTON ROAD	
		CITY: ROCHESTER	
		STATE: NEW YORK	
		COUNTRY: U.S.A.	
		ZIP: 14618	
		COMPUTER READABLE FORM:	
		MEDIUM TYPE: DISKETTE, 3.5 INCH	
		COMPUTER: IBM PC COMPATIBLE	
		OPERATING SYSTEM: MICROSOFT WINDOWS 95	
		SOFTWARE: MICROSOFT WORD VERSION 7.0A	
		CURRENT APPLICATION DATA:	
		APPLICATION NUMBER: US/09/297,928	
		FILING DATE: 11-MAY-1999	
		CLASSIFICATION: <unknown>	
		PRIOR APPLICATION DATA:	
		APPLICATION NUMBER: 60/03602	
		FILING DATE: NOVEMBER 13, 1996	
		ATTORNEY/AGENT INFORMATION:	
		NAME: FLOYD, LINDA AXAMETHY	
		REGISTRATION NUMBER: 33,692	
		REFERENCE/DOCKET NUMBER: CR-9981-P1	
		TELECOMMUNICATION INFORMATION:	
		TELEPHONE: 302-892-8112	
		TELEFAX: 302-773-0164	
		TELEX: 67117325	
		INFORMATION FOR SEQ ID NO: 15:	
		SEQUENCE CHARACTERISTICS:	
		LENGTH: 709 amino acids	
		TYPE: amino acid	
		STRANDEDNESS: unknown	
		TOPOLOGY: unknown	
		MOLECULE TYPE: protein	
		SEQUENCE DESCRIPTION: SEQ ID NO: 15:	
		US-09-297-928-15	
		Query Match	
		Best Local Similarity 5.9%; Score 167; DB 4; Length 709;	
		Matches 116; Conservative 66; Mismatches 182; Indels 188; Gaps 25	
OY	46	VQSOSQEIWAQVCISIRNA-LTLADVCAQSAV-----GIGFDATCSLVYLDKNG	93
DB	153	VECHQPKLLVNVAOCIASLSLQFINSESVANGLPKYVICMGIANMBRETIILMSRTFG	212
OY	94	DPLPVSPGDAAQNIIIVMDNRATE-----QAERIINNTHPRVLNVYGCKISP	140
DB	213	KPI-----VNCIVIMWDTITIVIRDKMONTVSVDROLRLRKGTGRPLL-----ST	257

```

0Y 141 EMEPKIAMLKENMP-----EIERGQGFDDLADFLTW-----RATGDLARSCTVTC 188
Db 258 YFSCCKLWFLDNELCTKAIENBDMFGTYD--TWLIYOLTKOKAIVSDVTANASRTGFM 315
0Y 189 KWTWLAHNR---WDPD-/------YFTIGLAELADEDEFIRIGH-----223
Db 316 NLSTLKYDNELLEFGIDKNLIMPEIYSSQYGGDEGI-----PDWIMEKLDHDSPKTVL 370
0Y 224 --IYSPGIFPCNGNLTAQAAALEMGLLPGRVA-----VGLIDMAGGIGIVGEGGLNN 275
0Y 276 LAYVGTSSCTMASTSPSFVGVWGPY---YSAMVPGI-----WLVGSGSAGAAALDQL 328
Db 431 LAF-----WEPHLOEYGGOKPELSPHFALGESVAVAGAVO-- 466
0Y 339 LDFHVAEAREMAQRVNQPLPWMLAD--RILEKTAOP--SDAVALAKGLHVPEFLGN 383
Db 468 -----WLRDLRLRIDKSEDDGPASTVDPDSGGVFEVPFSGI 504
0Y 384 RAPFADPHARAVICGLGMEERDLNLLALYINGLCIGIGYGLRQLIDAT--AQGVSN-- 433
Db 505 FAPYDPPARATIMGMSQFTTASHIARAAYGVC--FQAAILIKAMSSDAFEGSGKRD 561
0Y 440 -----IVISGAGQHPHYRQIILADTCGIPVI-----TTQCEPVLG 476
Db 562 FLEISDVTYERKSPSLVLAIVGSGMSRSENVQIQADILG-PCVAVRSPRAEC----- 613
0Y 477 SAILGAVAGNIA 488
Db 614 TALGAIAANMA 625

RESULT 14
US-09-036-987A-5
Sequence 5, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESSES:
ADDRESS: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: India
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847

```

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4928 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-036-987A-5

Query Match 3.8%; Score 106.5; DB 4; Length 4928;
 Best Local Similarity 23.8%; Pred. No. 1.7;
 Matches 126; Conservative 65; Mismatches 201; Indels 137; Gaps 31;

QY 34 EKITTTT-----GSRVEQS-----SQE--IWQAVCSIRNALTLADY 70
 Db 2031 ERLSDARRNGHEVLAVVRSANODGASNGLTAPNGPSQQRVITQALTSA---GLSVSDV 2087
 QY 71 CAQSVAGIG-----FDATCSLVYLDKNGDPLPVSPGDAKONIIYMDHRATQARIN 124
 Db 2088 DAVEAHGTGRLGDPTEAQAALIAATYGRDRPGRPLWLGVSXNT----- 2131
 QY 125 ATHHPLVNYGKISPEM-----ETPKILWKEKMEPEYERAGQFEDLADFLTRATGDL 179
 Db 2132 -GHTQAAGVAGYIKVMMAKRGELPRTLHYDEPSAOVMSAGTVOLLTENTPDPDSGRL 2190
 QY 180 ARS-VCTVTCKWTWLAH-----ENRMDPDYFRTIGLAELADEDFTRIGHIIVSPG 228
 Db 2191 RRAGVSSEFGISGT-NAHLILEQPPRESQSRSTEDP-----SGSVRDF-PVVPWNVSGK 2240
 QY 229 TPCGNGLTQAQAAEMGILLPCTPVAVGLIDAHAGIG-TVYEGGALNNLAVFGTSCTM 287
 Db 2241 TP--EALSAQADALMSYLSNR-----VDASPRDIGYSLAVTRPALDHRRAVVLGADRAAL 2292
 QY 288 -----ASTTSPSPVPGWGPYYSAMVPGLWLVGGO--SAGAAIDQLLD-FHPAVEA-- 338
 Db 2293 LPLKALAVSNDAAEVITGTRAAAGPVGFVSGGGQMPGMSGLHSAFPVFADAFDEACC 2352
 QY 339 -----REMAQRVNOPLPVWLAD-----RILEKT--AOPSDAVALAKGLHVVPEFLGN 383
 Db 2353 ELDAHLCQMAR-----LRDVLSSGSDPQLLDQTLMAQPG-LFALQVGLM--ELLGS 2399
 QY 384 RAPADPHARAIVICGLMERDLDNLALYTAGLCIGYGLRQIID-AQTAQGVSKNIVI 442
 Db 2400 WGV-----RPVVVLG-----HSVGEIAAFAAGVLSLRDAARLVAGRRLMQALPTGGAML 2450
 QY 443 SGGAGQHPVRLQIADTCGIPVITTCCEP---VLGSA-ILGAVAGNI 487
 Db 2451 AAAAGEBQL-RPLIAD-CGDRVGIAAVNAPGSVVLGSDRDVLDIDIGRL 2497

RESULT 15

US-09-370-700-5

; Sequence 5, Application US/09370700
 ; Patent No. 6274350
 ; GENERAL INFORMATION:
 ; APPLICANT: Baltz, Richard H
 ; APPLICANT: Broughton, Mary C
 ; APPLICANT: Crawford, Kathryn P
 ; APPLICANT: Madduri, Krishnamurthy
 ; APPLICANT: Treadway, Patti J
 ; APPLICANT: Turner, Jan R
 ; APPLICANT: Waldron, Clive
 ; TITLE OF INVENTION: Biosynthetic Genes for Spinosyn Insecticide
 ; FILE REFERENCE: 50489 DIV1
 ; CURRENT APPLICATION NUMBER: US/09/370,700
 ; CURRENT FILING DATE: 1999-08-09
 ; EARLIER APPLICATION NUMBER: US 09/36987
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 4928
 ; TYPE: PRT
 ; ORGANISM: Saccharopolyspora spinosa

US-09-370-700-5

Query Match 3.8%; Score 106.5; DB 4; Length 4928;
 Best Local Similarity 23.8%; Pred. No. 1.7;
 Matches 126; Conservative 65; Mismatches 201; Indels 137; Gaps 31;

QY 34 EKITTTT-----GSRVEQS-----SQE--IWQAVCSIRNALTLADY 70
 Db 2031 ERLSDARRNGHEVLAVVRSANODGASNGLTAPNGPSQQRVITQALTSA---GLSVSDV 2087
 QY 71 CAQSVAGIG-----FDATCSLVYLDKNGDPLPVSPGDAKONIIYMDHRATQARIN 124
 Db 2088 DAVEAHGTGRLGDPTEAQAALIAATYGRDRPGRPLWLGVSXNT----- 2131
 QY 125 ATHHPLVNYGKISPEM-----ETPKILWKEKMEPEYERAGQFEDLADFLTRATGDL 179
 Db 2132 -GHTQAAGVAGYIKVMMAKRGELPRTLHYDEPSAOVMSAGTVOLLTENTPDPDSGRL 2190
 QY 180 ARS-VCTVTCKWTWLAH-----ENRMDPDYFRTIGLAELADEDFTRIGHIIVSPG 228
 Db 2191 RRAGVSSEFGISGT-NAHLILEQPPRESQSRSTEDP-----SGSVRDF-PVVPWNVSGK 2240
 QY 229 TPCGNGLTQAQAAEMGILLPCTPVAVGLIDAHAGIG-TVYEGGALNNLAVFGTSCTM 287
 Db 2241 TP--EALSAQADALMSYLSNR-----VDASPRDIGYSLAVTRPALDHRRAVVLGADRAAL 2292
 QY 288 -----ASTTSPSPVPGWGPYYSAMVPGLWLVGGO--SAGAAIDQLLD-FHPAVEA-- 338
 Db 2293 LPLKALAVSNDAAEVITGTRAAAGPVGFVSGGGQMPGMSGLHSAFPVFADAFDEACC 2352
 QY 339 -----REMAQRVNOPLPVWLAD-----RILEKT--AOPSDAVALAKGLHVVPEFLGN 383
 Db 2353 ELDAHLCQMAR-----LRDVLSSGSDPQLLDQTLMAQPG-LFALQVGLM--ELLGS 2399
 QY 384 RAPADPHARAIVICGLMERDLDNLALYTAGLCIGYGLRQIID-AQTAQGVSKNIVI 442
 Db 2400 WGV-----RPVVVLG-----HSVGEIAAFAAGVLSLRDAARLVAGRRLMQALPTGGAML 2450
 QY 443 SGGAGQHPVRLQIADTCGIPVITTCCEP---VLGSA-ILGAVAGNI 487
 Db 2451 AAAAGEBQL-RPLIAD-CGDRVGIAAVNAPGSVVLGSDRDVLDIDIGRL 2497

Search completed: March 13, 2003, 16:42:17
 Job time : 21.5877 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:12 ; Search time 15.0298 Seconds

(without alignments)
1637.624 Million cell updates/sec

Title: US-09-802-208b-4

Perfect score: 2811
Sequence: 1 MTTKTVIGVDVSGSVRAG.....HRRYEVAYKQLOHTAKLLRD 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications-AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2811	100.0	534	9	US-09-802-208b-4
2	306	10.9	484	10	US-09-815-242-10383
3	303	10.8	501	10	US-09-815-242-10866
4	302.5	10.8	503	10	US-09-815-242-11080
5	289	10.3	517	10	US-09-815-242-12837
6	288	10.2	509	9	US-09-815-242-5796
7	279.5	9.9	509	9	US-09-738-626-6686
8	276.5	9.8	502	10	US-09-815-242-10425
9	276.5	9.8	502	10	US-09-815-242-11431
10	273.5	9.7	502	10	US-09-815-242-13660
11	267.5	9.5	505	9	US-10-116-821-14
12	264.5	9.4	496	10	US-09-815-242-5321
13	264.5	9.4	498	10	US-09-815-242-12335
14	262.5	9.3	501	10	US-09-815-242-13788
15	254	9.0	505	10	US-09-815-242-11927
16	246.5	8.8	496	10	US-09-815-242-10714
17	221	7.9	473	9	US-09-738-626-3630
18	219.5	7.8	447	10	US-09-815-242-11747
19	187.5	6.7	494	9	US-09-738-626-6829

20	167	5.9	709	9	US-09-308-207-18	Sequence 18, Appl
21	164	5.8	518	12	US-10-001-852-12	Sequence 12, Appl
22	108.5	3.9	513	10	US-09-833-745-63	Sequence 63, Appl
23	107	3.8	439	9	US-09-820-843A-60	Sequence 60, Appl
24	104.5	3.7	527	9	US-10-176-640-3	Sequence 3, Appl1
25	101.5	3.6	63	10	US-09-764-847-580	Sequence 580, App
26	101.5	3.6	650	10	US-09-815-242-13341	Sequence 13341, A
27	101.5	3.6	2993	9	US-09-738-626-6239	Sequence 6239, Ap
28	101	3.6	943	10	US-09-815-242-12027	Sequence 12027, A
29	100.5	3.6	438	9	US-09-738-626-4285	Sequence 4295, Ap
30	100.5	3.6	729	10	US-09-287-849-2	Sequence 2, Appl
31	99	3.5	881	10	US-09-838-539-8	Sequence 8, Appl1
32	97.5	3.5	570	9	US-09-738-626-5603	Sequence 5603, Ap
33	97.5	3.5	3739	9	US-09-860-846-33	Sequence 33, Appl
34	97.5	3.5	3739	10	US-09-860-846-6	Sequence 6, Appl1
35	97.5	3.5	11877	9	US-09-861-289-33	Sequence 6, Appl1
36	97.5	3.5	11877	10	US-09-861-289-6	Sequence 32, Appl
37	97	3.5	701	10	US-09-900-237-32	Sequence 6, Appl1
38	96.5	3.4	403	10	US-09-996-194-6	Sequence 10785, A
39	96.5	3.4	640	10	US-09-815-242-10786	Sequence 261, App
40	95.5	3.4	1827	9	US-09-712-363-261	Sequence 15, Appl
41	95.5	3.4	4999	9	US-09-976-059-15	Sequence 20, Appl
42	93.5	3.3	506	10	US-09-900-237-20	Sequence 2, Appl1
43	93	3.3	1060	10	US-09-955-909-2	Sequence 37, Appl
44	92.5	3.3	740	9	US-10-051-909-37	Sequence 30, Appl
45	92.5	3.3	1080	10	US-09-900-237-30	

ALIGNMENTS

RESULT 1

US-09-802-208b-4
; Sequence 4, Appl1
; Publication No. US20030041352A1
; GENERAL INFORMATION:
; APPLICANT: Parrot, Wayne
; APPLICANT: Lafayette, Peter
; APPLICANT: Kane, Patrick
; TITLE OF INVENTION: Antibiot or Ribitol As Positive Selectable Markers
; FILE REFERENCE: UGA-855R
; CURRENT APPLICATION NUMBER: US/09/802,208B
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-802-208b-4

Query Match	Score	2811	DB 9	Length	534
Best Local Similarity	100.0%		Pred. No. 7.1e-251		
Matches 534	Conservative	0	Mismatches	0	Indels
0	Gaps	0			
OY	1	MTTKTVIGVDVSGSVRAGIFDNGSLSHATEKITTRRSGRVROSOEIMQAVCSC	60		
DB	1	MTTKTVIGVDVSGSVRAGIFDNGSLSHATEKITTRRSGRVROSOEIMQAVCSC	60		
OY	61	IRNALTLADYCAQSVAGIGDATCSLVLDKNGDPLFVSPGDAKONIYMDHRAEQ	120		
DB	61	IRNALTLADYCAQSVAGIGDATCSLVLDKNGDPLFVSPGDAKONIYMDHRAEQ	120		
OY	121	ERINATHHPVLNVYGGISPEMETPKILMKENNPETERYERAGOFFDLADFLVTRATCDLA	180		
DB	121	ERINATHHPVLNVYGGISPEMETPKILMKENNPETERYERAGOFFDLADFLVTRATCDLA	180		
OY	181	RSVCTVCKMTWLAHERMPDYFRITGLADEDEIRIGHIHVSPTCGNGLTROAA	240		
DB	181	RSVCTVCKMTWLAHERMPDYFRITGLADEDEIRIGHIHVSPTCGNGLTROAA	240		
OY	241	AEMGLLPGRVAVAGLIDAHAGIGTGVVEGALNNLAIVFGTSCTMASTTSPFVPGVW	300		

```
Db 241 AEMGLGCTPFAVGLIDAHAGCTVGEVAGALNNLAIVETSSCTMASTTSPSFEVGVW 300
Qy 301 GPPYSAMVPGMLVEGQSAGAAIDQLDFHPAVEAREMAQVNOPLPWLADRILEK 360
Db 301 GPPYSAMVPGMLVEGQSAGAAIDQLDFHPAVEAREMAQVNOPLPWLADRILEK 360
Qy 361 TAQSSDAVALAKGLHVPPEFLGNRAPADPHARAVICGLCHERLDNLALYINGLCIG 420
Db 361 TAQSSDAVALAKGLHVPPEFLGNRAPADPHARAVICGLCHERLDNLALYINGLCIG 420
Qy 421 YGLROIIDAQTAQGVSKNIYISGAGOHPLVROIADTCIPYITTOCCPEVLLGSAIL 480
Db 421 YGLROIIDAQTAQGVSKNIYISGAGOHPLVROIADTCIPYITTOCCPEVLLGSAIL 480
Qy 481 GAVAGNIAPSVGEAMQOFTHVDKYYRPOERYQSLHHRREAYAKOLQHTAKLLRD 534
Db 481 GAVAGNIAPSVGEAMQOFTHVDKYYRPOERYQSLHHRREAYAKOLQHTAKLLRD 534

RESULT 2
US-09-815-242-10383
; Sequence 10383, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10383
; LENGTH: 484
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-815-242-10383

Query Match 10.9%; Score 306; DB 10; Length 484;
Best Local Similarity 24.4%; Pred. No. 6.1e-20;
Matches 132; Conservative 75; Mismatches 249; Indels 86; Gaps 17;
```

```
Db 106 TLEEARVPOSQVITGNLMMPGEFTAPKLLMVOHREPEIFRQIDKVLKROYLRLMTGEFA 165
Qy 181 RSVCTVTCCKMTWLAHENR-MDPDFERTIGLAELADEDFIRIGHIVSPGTPCGNLTAQA 239
Db 166 -SDMSDAAGTMTLVNARNDSDVLMQAC-----DLRQMPALVEGSRITGALLPEV 216
Qy 240 AAEMGLLPGTFVAVGLIDAHAGGIGTVGEGALNNLAIVFGTSSCTMASTTSPSPVGV 299
Db 217 AKAGMAT-VFVVVGGGDMNAGAVGVMDA---NQAMLSTGSGVYFA--VSEGFLSKP 270
Qy 300 WGPYYS--AMVPGMLVEGQSAGAAIDQL-----LDHPAVEAREMAQVNOPLPWL 352
Db 271 ESAVHSCFALPQRNHLMSWLSASCLDMAKLTGSLNVPALAAQAQADESNE--PVM 328
Qy 353 LADRILEKTAQPSDAVALAKGLHVPPEFLGNRAPADPHARAVICGLCHERLDNLALY 412
Db 329 F-----LPTLSGERTEHNNPQAKGVFEGLTHQGPHELAR-- 363
Qy 413 IAGLCGIGYGLROIIDAQTAQGVSKNIYISGAGOHPLVROIADTCG--IPYITTOCC 470
Db 364 -AVLEGVYALADMDVNHACGIRPQSVTLIGGARSEYWRQMLADISGQOLDYRTGDV 422
Qy 471 EPVLLGSAILGAVAGNIAPSVGEAMQOFTHVDKYYRPOERYQSLHHRREAYAKOLQHTAK 530
Db 423 GPA-LGAARLAQIAPANEKSLIELLPOL-----PLEGSHLPDAQRYAAYQPRRETFR 473
Qy 531 LL 532
Db 474 RL 475
```

```
RESULT 3
US-09-815-242-10866
; Sequence 10866, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10866
; LENGTH: 501
; TYPE: PRF
; ORGANISM: Enterococcus faecalis
US-09-815-242-10866

Query Match 10.8%; Score 303; DB 10; Length 501;
Best Local Similarity 23.4%; Pred. No. 1.2e-19;
```

Matches 132; Conservative 82; Mismatches 235; Indels 114; Gaps 21;

```

Oy 1 MTITKTVIGDVGGSSVRAGIFDNLGSLSHATEKTTTRRSGSRVQSSOEIOWAVCGC 60
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 1 MAEEYIYMAIDGTTSSRAIIFDKGNKIGSSOKEFTQYFPNAGVNEHANEIWNVSQSV 60
Oy 61 IRNALTLADVCAOSVAGIFDNLGSLSHATEKTTTRRSGSRVQSSOEIOWAVCGC 120
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 61 IAGSLSESVKPTDIAGITITNORETIVYWDKATGPT-----YNALVWOSRQTTPIA 113
Oy 121 ERINATNHPVLY--VGGKISPEMETPKILMLKENNPEIYERAG---OFFDLADELTWR 174
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 114 DOLKEDGYSEMIHEKTLIIADYFSAATKYRMILDHVEGAQERANGELMEGTIDTFLVWK 173
Oy 175 ATGDL-ARVCTVYCKMTLHLENRMDPRYFRIGLAELADEDFIRIGHIYSPGTPCG 233
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 174 LUGDHHVTDYTNASRTMLNIDLDMDDELMLNLTIP-----RVMLPKVVSNS 221
Oy 234 ---GLTAQ---AAAEKGLPPTPVAVGLIDAHAGIGTVEGALNMLAYVEGTSSCTM 287
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 222 EYVGLTKNHFYGSFV-----PIAGMADQQAALFGQMAFEGMWKN---TYGTGSFTY 272
Oy 288 AST-----TSPSFVPQWGPYVSAMVPGMLVEGQSAAGAAIDOLLDFHPAV 335
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 273 MMTGEPPOLSKNNLTITIGY--GINKGYVAL-----EGSIFVAGSAIQ----- 314
Oy 336 EEARERAAQVNOPLPYWLADRI-LEKTAOPSDAVNALAKGLH---VYPEFLGNRPFADP 390
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 315 -----WLDGKMLQOTAAESEVAANASGHNVEVYVPAFTGAGAPYWD 358
Oy 391 HARAIVGIGMERDLNLTALYIAGLCIGYLRQIILDAQTAQ--GVASNIYVSGAGOH 449
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 359 QARGAVFGILRTGRTREDFK---ATIQAVAYQROLDITMKKEPTGIDIVLAKYDGAANN 415
Oy 450 PLVROILADTCGIPVTTTCCCEPVLLGSAIILGAVAGNIAPSVG-----EAMQOFTHVXY 504
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 416 DELMOPQADILNTAVGRAHNLLETTALGAALAGLA-----VGFMDLEIKAFQEEGOQ 469
Oy 505 YVP---QERYOSLHRRYEAYKO 524
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 470 FEPYMAEEREDL---YEGWOD 488

```

RESULT 4
US-09-815-242-11080
; Sequence 11080, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11080
; LENGTH: 503
; TYPE: PR
; ORGANISM: Haemophilus influenzae
US-09-815-242-11080

Query Match 10.8%; Score 302.5; DB 10; Length 503;
Best Local Similarity 23.3%; Pred. No. 1.4e-19;
Matches 128; Conservative 76; Mismatches 259; Indels 87; Gaps 18;

```

Oy 1 MTITKTVIGDVGGSSVRAGIFDNLGSLSHATEKTTTRRSGSRVQSSOEIOWAVCGC 60
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 1 MTDKTYIADLDOGTSSRAVLIDHNNANVEIAREFTQIYPRAGVNEHMEIATQOST 60
Oy 61 IRNALTLADVCAOSVAGIFDNLGSLSHATEKTTTRRSGSRVQSSOEIOWAVCGC 120
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 61 LNEVAKAGITSDLEIAITITNORETIVYWDKATGPT-----YNALVWOSRQTTADI 113
Oy 121 ERINATNHP--VLNVYGGKISPEMETPKILMLKENNPEIYERAGOFFDLAD-FLTW 174
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 114 DKLKADGHEIYIRNTGLVADPYFSGTQYKWMILDNVEGAREKAREGELLFGTVDTLVWK 173
Oy 175 ATGDLARVCTVYCKMTL--AHENRMDPRYFRIGLAELADEDFIRIGHIYSPGTPCG 232
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 174 LUGRHHVTDYTNASRTMLNIDLDMDDELMLNLTIP-----RVMLPKVVSNS 221
Oy 233 NGILTAQAAEMGLPPTPVAVGLI--DAHAGIGTVEGALNMLAYVEGTSSCTMAST 290
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 233 KG-----GVRIPVAGIADQQAALGHLCHVHAGAKN---TYGT-GCFMLH 275
Oy 291 TSPSFVPQWVG-----PYVSAMVPGMLVEGQSAAGAAIDOLLDFHPAVEA 338
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 276 TGNKATTSKNGILTTIACNAKGEPEYA-----LEGSVFIAGASIOMLRDLKTVH 327
Oy 339 RE---MAORVNOPLPYWLADRILEKTAOPSDAVNALAKGLHVEPEFLGNRPFADP 395
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 328 FDEYFRAQVTD-----SNGYIVPAFTGLGAPYWDYARKA 364
Oy 396 ICGIGMERDLNLTALYIAGLCIGYLRQIILDAQTAQ--GVASNIYVSGAGOH 454
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 365 IFTLSGARNNHIVR---ATLESIAVQTRVLEAMQSDSEGRQYLRVQDAGANNFLMQ 421
Oy 455 ILADTCGIPVTTTCCCEPVLLGSAIILGAVAGNIAPSVGAMQOFTHVXYQERYOSL 514
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 422 FQADILDVNERPVYKREVTAALGAAYLAGLATGFWKDLDE-LRDKARVERTFSPDSNEK 479
Oy 515 HRRYEAYKO 524
    | : : : : | : | : | : : : : : : : : : : : : : : : : : : :
Db 480 RRRYKMKK 489

```

RESULT 5
US-09-815-242-12837
; Sequence 12837, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12837
LENGTH: 517
TYPE: PRF
ORGANISM: *Staphylococcus aureus*
US-09-815-242-12837

Query Match 10.3%; Score 289; DB 10; Length 517;

Best Local Similarity 22.9%; Pred. No. 2,5e-18; Matches 131; Conservative 92; Mismatches 244; Indels 104; Gaps 23;

QY 5 KTVIGVDSGSRVAGIFDLNGS-LLSHA-----TEKITTRSGSRVSGSOEIMQAV 57
DB 2 KYMIGVDIGTSTKSVLYDENGTFIMKHQIGYDLHTPNVDVS-----EENPELDFAV 54
QY 58 CSCIRNMLTLADYCAQSVAGIGFDATC-SLYVLDKNGDPLPVSPGDAKQNIYVMDHRA 116
DB 55 LMTIKYVRESKVKKDDIKFVSFSQMSHSLAMDQOHR-----TNNITWADNRA 105
QY 117 TEQAEIRINATH--HPVLNVGKISPEMETPKILMKENPEIYERAGOFPLADFLTW 174
DB 106 AKYATVINEVDGNATYQRTGTPRHPMSPLAKIFMKHEMDVQRTAKAFDITTYIF 165
QY 175 A--TGDLARVCTVTCWKWTLAHENRWDPDYFRTIGLA-ELADEDFIRIGHIIVSPGTPC 231
DB 166 LFDYVYIIDSMSASATGMFNLETLID--MDVEALELIGISKEMLP-----LVPTTYV 214
QY 232 GNGTLAQAAMGLPGTPAVANGLIDAHAGIGTVGEGGLNNLAVFGSSCTMASTT 291
DB 215 MKGKERRATLMLGINKDTPYVIGASD--GVLNLSGVNSVKGGEVAATVIGSGAIRTV 271
QY 292 SP-----SFVPGVWGPPYYSAMVPGMLVVEGSGSAGAALDOLDLDFHRAV 335
DB 272 KPRTDYGRIFCYVLAEDHYIG--GPVNNGVVLRML-----RDELL--ASEV 316
QY 336 EAREMAQORVNOPLPVMLADRILEKTAQPSDAVALAKGLHVPFLGNRAPFADPHARAV 395
DB 317 ETARRLV---DPYDVLQ---IAKRVKPG-----ADGLIFHPYLAGERAPLMMANAR 365
QY 396 ICGGMRBDNLALTYAGLCIGYGLROI-LDAQTAQGVVSKNIYISGAGOHPLVRQ 454
DB 366 FFGTLTHKKEHMR---AALEGVLYNLYTYVLLIEVNMETPMKATGGAFASEVW 422
QY 455 ILADTCGIPVITTOCEPVLGSAIILGAVA-----GNIPASVEGAQOEFHNVK 504
DB 423 MMSDIFTELVPESYESSCGACVGLKAVGDIEDFSIYSSMGATNNHPIEENTVY 482
QY 505 YFPOERQSLHHRREYAKQI-----OHTAK 530
DB 483 QETVSIPIINLSRSLTENYEQIADFOROHIAE 513

RESULT 6
US-09-815-242-5796
Sequence 5796, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5796
LENGTH: 509
TYPE: PRF
ORGANISM: *Staphylococcus aureus*
US-09-815-242-5796

Query Match 10.2%; Score 288; DB 10; Length 509;

Best Local Similarity 23.6%; Pred. No. 3e-18; Matches 122; Conservative 83; Mismatches 221; Indels 90; Gaps 21;

QY 3 ITKTIVGVDSGSRVAGIFDLNGS-LLSHA-----TEKITTRSGSRVSGSOEIMQ 55
DB 1 IMKYMIGVDIGTSTKSVLYDENGTFIMKHQIGYDLHTPNVDVS-----EENPELDF 53
QY 56 AVCSICRNALTLADYCAQSVAGIGFDATC-SLYVLDKNGDPLPVSPGDAKQNIYVMDH 114
DB 54 AVLMTIKYVRESKVKKDDIKFVSFSQMSHSLAMDQOHR-----TNNITWADN 104
QY 115 RATEQAEIRINATH--HPVLNVGKISPEMETPKILMKENPEIYERAGOFPLADFLT 172
DB 105 RAAKYATVINEVDGNATYQRTGTPRHPMSPLAKIFMKHEMDVQRTAKAFDITTYIF 164
QY 173 WRA--TGDLARVCTVTCWKWTLAHENRWDPDYFRTIGLA-ELADEDFIRIGHIIVSPG 229
DB 165 YHLDYVYIIDSMSASATGMFNLETLID--MDVEALELIGISKEMLP-----LVPTT 213
QY 230 PCGNGTLAQAAMGLPGTPAVANGLIDAHAGIGTVGEGGLNNLAVYFGSSCTMAS 289
DB 214 YVMGKERRATLMLGINKDTPYVIGASD--GVLNLSGVNSVKGGEVAATVIGSGAIRTV 270
QY 290 TTSP-----SFVPGVWGPPYYSAMVPGMLVVEGSGSAGAALDOLDLDFH 333
DB 271 IDRRTDYGRIFCYVLAEDHYIG--GPVNNGVVLRML-----RDELL--AS 315
QY 334 AVEAREMAQORVNOPLPVMLADRILEKTAQPSDAVALAKGLHVPFLGNRAPFADPHAR 393
DB 316 EYETAKRLGV---DPYDVLQ---IAKRVKPG-----ADGLIFHPYLAGERAPLMMANAR 364
QY 394 AVTICGMRBDNLALTYAGLCIGYGLROI-LDAQTAQGVVSKNIYISGAGOHPLV 452
DB 365 GSEFGTLTHKKEHMR---AALEGVLYNLYTYVLLIEVNMETPMKATGGAFASEV 421
QY 453 ROLLADTCGIPVITTOCEPVLGSAIILGAVA--GNI 487
DB 422 ROMMSDIFTELVPESYESSCGACVGLKAVGDI 457

Db 273 LMTGEEAVSENGSLTTACGPTGEVN---YALGAVFNAGASIQ----- 315
Qy 340 EMAQRVNPPLPWLADRI-LEKTAQPSDAVAL---AKGLHVVEFLGNRAPFADPHARA 394
Db 316 -----MLRDEMKLINAYDSEYFATKYQNTNGVYVAFGLGAPYWDYPARG 363
Qy 395 VIGGLGHERDLNLAALYIAGLCIGYGLROIIDA-OTAGGVYSKNIVISGAGQHPYLR 453
Db 364 AIFGLTRGVANNAHRII---ATLESIAVQTRVLEAMQADSGIRLHALRVGGAVANFELM 420
Qy 454 QILADTGPVITTOCEPVLGSAIIGAVAGNAPSVGEAMQOFTVDKYTP-----QE 509
Db 421 QFOSDILGTVREVEREVATLGAAYLAGLAVGFQWNLDE-LOEKAVIEREFPGIETTE 479
Qy 510 RYOSLHRRRYEAYKQ 524
Db 480 R-----NYRYAGWK 489

RESULT 9
US-09-815-242-13431
; Sequence 13431, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13431

Query Match 9.8%; Score 276.5; DB 10; Length 502;
Best Local Similarity 24.4%; Pred. No. 3.4e-17;
Matches 140; Conservative 83; Mismatches 235; Indels 115; Gaps 23;

Qy 1 MTTTIGVGVGSGVRAGIFDLNGLSHATEKITTRSGSRVDSOSSOEIQAQASC 60
Db 1 MSQEKYIMADQGTTSRAIIFNNKSGEKVSQKEFTQIIPQAGVHEHNNNEIINSQSV 60
Qy 61 IRNALTLADVCASVAGIGF-DATCSLVLDKNGDPLVPSPBGDAKONITVMDHRTAQ 119
Db 61 IAGAFIESGVKPNQIEAIGITNQRETTVVDKK-TGLPI-----YNATVMSRQTAPL 112
Qy 120 AERINATHHVLNV-----GKISPEMETPKILMLKNNMPEIYERAO-----FFDLA 168

Db 113 AEOLKSQ-----GYEFHEKTEGLIIDAYFSATYKVRMILDHVEGAQRAEKEGELLFCTID 167
Qy 169 DELTWRAATGLDARSVCVTCTCKWTLAH--ENRMPDVFRTIGLAELADEDFIRIGHIYS 226
Db 168 TWLVWKTLIDGAHAHTDVSNAARTMLYNIKELKMDLEILEINIKALPE-VRSNSLEY 225
Qy 227 PGTFCNGNLTAQAABEGLDGPVAVGLIDAHAGIGTVGVEGALNNLAVFGTSSCT 286
Db 226 -----GKTAPHFHFYGEYVDSGMAG---DOQAALFGOLAFEBGMWKN--TYGTGSFI 272
Qy 287 MAST-----TSPSPVPGWMPYYSAMVGLMVEGGGSAAGAATDOLDLHPA 334
Db 273 IMNTGEEMOSENNLLTITG--GINKGVYAL-----EGSFTIGSALQWLRDGLRM 323
Qy 335 VEEAREMAQRVNPPLPWLADRILEKTAOPS---DAVALAKGLHVVEFLGNRAPFADPH 391
Db 324 VENSPE-----EKYARBSHNNDEV-----YVPAFGLGAPYWNQN 360
Qy 392 ARAVTCGLGHERDLNLAALYIAGLCIGYGLROIIDA-OTAGGVYSKNIVISGGA 446
Db 361 ARGSVFGLTRGTSKEDEIK--ATLQSIAYQVRDIDTMOVDTOTAIQVLK---VDGGA 413
Qy 447 GQHPVLROIADTGPVITTOCEPVLGSAIIGAVAGNAPSVG-----EAMQOFTHV 501
Db 414 AMNFFLMQFOADILGIDIAKKNLETTALGAAFLAGL-----SVGYWKDLDELKLNET 467
Qy 502 DKYVYP---QERYOSLHRRRYEAYKQLOHTRAKL 531
Db 468 GELFEPMSNESRKQOLYKGMKAVKATQVFAEV 500

RESULT 10
US-09-815-242-13660
; Sequence 13660, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13660
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13660

Query Match 9.7%; Score 273.5; DB 10; Length 502;
Best Local Similarity 24.4%; Pred. No. 6.4e-17;

	Matches	140:	Conservative	82:	Mismatches	236:	Indels	115:	Gaps	23:
OY	1	MTITRTVIGVDVSGSVRAGIFPDLNGSLLSHATEKITTTTRSGSVNEOSSOEIWOAVCSG	60							
	:	: :								
Db	1	MSOEKYIMAIIDOGFTTSRAIIIFNNKKGKEVSSSQKEFTQIFPGACVVEHNANEINVSQSV	60							
OY	61	IRNALTLADVCAGSVAIGGF-DATCSLVVLDDKNBPRLVSPBGDAKONITYMMDHRATEQ	119							
	:	: :								
Db	61	IAGAFIEGSVPKPNDEAIGITNORETTYVMDBKK-TGLPI-----YNAIWQSROTAPL	112							
OY	120	AERINATHPLVNLV-----GGKISPEMEPTKIYMKENMPLEYERAGO----FPDLA	168							
	:	: :								
Db	113	AEOLKNO-----GYEKHEKHTGLIIDAYFSATKVKMILLDHVEGAOEREKELLEFTID	167							
OY	169	DELTPRATGDILARSCTYTCKWTWLAH--ENRMDPDYFRTIGLALADEDFIRIGHNIYS	226							
	:	: :								
Db	168	TWLWVKLLDGAANHVDYSNARTMLYNIKELKMWDEILELINPKAILPE-VRSNSEIY-	225							
OY	227	PCTPCNGCLTAOAAEMKLLPETPYAVGLIDAHAGSIGTVGEVGALNNLAYVFSTSCOT	286							
	:	: :								
Db	226	-----GKTAPHFHYGSEVPISGMAG---DQOALLFOALFEPMVKVN---TYGTGSPFI	272							
OY	287	MAST-----TSPSFVPGWGMYTYSAMVGCGLVLEBGSQAAGAIDLQLDFHPA	334							
	:	: :								
Db	273	IMNGEEMLOSENMLTTIGY--GINGGVYYAL-----ESGIPIASGAIOMLRDGIRM	323							
OY	335	VEEAAREMAQRVNQPLRPVWLADRILETKTOPS---DAVALAKLHVYPFLGNRAPFDPH	391							
	:	: :								
Db	324	VENSES-----EKYARDSHINNDENV-----YVPAFTGICGAPYMWN	360							
OY	392	ARAVICGIGMERDLDDLNLALYIAGLCIGIGYGRQLT-----DAQTAQGVSKNIYISGA	446							
	:	: :								
Db	361	ARGSVFGLTGRGSKDKFLK--ATLQSIATQVRDIIDIMQVDTQAINIYLK----VDGA	413							
OY	447	GOHLPRROILLADTCGIPVITTOCCCEPVLLGSAILGAVAGNTAPSVG-----EAMQQTFHV	501							
	:	: :								
Db	414	AMNNFMLOFQADILDICIDIDARAKNMETTLALGAFFLAGL-----SYGWXKDIDELKLNET	467							
OY	502	DKYUYP--OERYOSLHHRRYEANYQLOHTAKL	531							
	:	: :								
Db	468	GELPEPSNESKRQEOLYRGWKKAIVATQVFAEV	500							
	RESULT 11									
	US-10-116-821-14									
	Sequence 14, Application US/10116821									
	Publication No. US20030040086A1									
	GENERAL INFORMATION:									
	APPLICANT: Dodge, Timothy C.									
	APPLICANT: Valle, Fernando									
	TITLE OF INVENTION: Methods for the Production of Products									
	TITLE OF INVENTION: In Host Cells									
	FILE REFERENCE: GC620-3									
	CURRENT APPLICATION NUMBER: US/10/116, 821									
	CURRENT FILING DATE: 2002-04-05									
	PRIOR APPLICATION NUMBER: US 60/282,277									
	PRIOR FILING DATE: 2001-04-05									
	NUMBER OF SEQ ID NOS: 14									
	SOFTWARE: FastSeq for Windows Version 4.0									
	SEQ ID NO 14									
	LENGTH: 505									
	TYPE: PRT									
	ORGANISM: Pantoea citrea									
	US-10-116-821-14									

Query Match 9.5%; Score 267.5; DB 9; Length 505;
Best Local Similarity 22.1%; Pred No. 2; 3e-16;
Matches 120; Conservative 90; Mismatches 25; Indels 79; Gaps 20.

OY 5 KTVICVDVGGSGVRAGCIDLNGSLSHATEKTTTTRSSGRVEQSSQETIQAVCSCIRNA 64
Db 7 KYVALDGTSTSRVVLDHDNANILAVSRETOHYHPTKGVEHDPMDIMAAQSSTLYEV 66

[illegible]

```

1  RESULT 12
2  US-09-815-242-5321
3  Sequence 5321, Application US/09815242
4  Patent No. US2002006156A1
5  GENERAL INFORMATION:
6  APPLICANT: Haselbeck, Robert
7  APPLICANT: Ohlsen, Karl L.
8  APPLICANT: Zyskind, Judith W.
9  APPLICANT: Wall, Daniel
10 APPLICANT: Trawick, John D.
11 APPLICANT: Carr, Grant J.
12 APPLICANT: Yamamoto, Robert T.
13 APPLICANT: Xu, H. Howard
14 TITLE OF INVENTION: Identification of Essential Genes in
15 TITLE OF INVENTION: Prokaryotes
16 FILE REFERENCE: ELTRA.011A
17 CURRENT APPLICATION NUMBER: US/09/815,242
18 CURRENT FILING DATE: 2001-03-21
19 PRIOR APPLICATION NUMBER: 60/191,078
20 PRIOR FILING DATE: 2000-03-21
21 PRIOR APPLICATION NUMBER: 60/206,848
22 PRIOR FILING DATE: 2000-05-23
23 PRIOR APPLICATION NUMBER: 60/207,727
24 PRIOR FILING DATE: 2000-05-26
25 PRIOR APPLICATION NUMBER: 60/242,578
26 PRIOR FILING DATE: 2000-10-23
27 PRIOR APPLICATION NUMBER: 60/253,625
28 PRIOR FILING DATE: 2000-11-27
29 PRIOR APPLICATION NUMBER: 60/257,931
30 PRIOR FILING DATE: 2000-12-22
31 PRIOR APPLICATION NUMBER: 60/269,308
32 PRIOR FILING DATE: 2001-02-16
33 NUMBER OF SEQ. ID NOS: 14110
34 SOFTWARE: FastSeq. for Windows Version 4.0
35 SEQ ID NO 5321
36 LENGTH: 496

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5321
Query Match
  9.4%: Score 264.5; DB 10; Length 496;
  Best Local Similarity 21.9%: Pred. No. 4.3e-16;
  Matches 112; Conservative 77; Mismatches 227; Indels 95; Gaps 17;

OY 5 KTVIGVDVSGSVRAGIFDLNGLSLSHATEKITTRRSRSGSVEQSSOEIMQAVCSIRNA 64
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 3 KYLISIDOGTSSRAILFNQKEIAGVAQREFKQYFQSGVNEHDANEIMTSVLAVMTEV 62
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 65 LTLADVCAQSVAGIGF-DATCSLVYLDKN-GDPLVSPREGDAKNIITVMDHRTGEQAE 122
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 63 INENDVRADQIAGIGITNORETIVVMDKHGTRPI-----YHAIWQSRQTSICSE 113
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 123 I--NATHNPVLNVYGKISPEMETPKILMLKENNPEIYERAGQ----FFDLADFLTRAT 176
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 114 LKQGYGQTFERDKTGLLDLPYFAGTKVKMLDNVEGAREKANGDILFGTIDTWLVKLS 173
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 177 GDLARSVCTVCKWTWL--AHENRMDPYERTIG----LAELADEDFI--RIGHIIVS 226
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 174 GKAAHITDYSNASTLMFNHIDEMDELLELLTPKRMLEPVKPSSEVYKTIIDYHFG 233
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 227 PGPFCGN-----GLTQAQAEMGILGPFPVAVGLIDAHAGIGITVGEAGALNLAIV 279
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 234 QEVPFIAGVAGDQALRGACFERGDVKNYGTGFMIMNTGD--KAVSESGLTTTAY- 291
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 280 FGTSSCTMASTTSPFVPGVWPYYSAMVPGMLVEGQSAAGAIDQLDFHPAVEAR 339
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 292 -----GIDGKYNVAL-----ESIFVSGSAIQ----- 313
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 340 EMAORVNPPLPWLADRI-LEKTAQPSDAVAL---AKGLHVEPELGNRAPDPHARA 394
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 314 -----WLRDGLRMINSAPQSESYATRVDSIEGVVPAFGLGIPYDSEARG 361
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 395 VICGLGMRDLNLLAIYIAGLCIGYGLRQILDQTA--QGVSKNIYISGAGQHPVLR 453
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 362 AIFGLTGTGTEKEHFIKATLESIC---YQTRDYMAMSKDSIDVQSLRVDGAVKNNFIM 418
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 454 QILADTCGIPYITTCCEPVLLGSAILGAVA 484
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 419 QFOADIYNTSVREPEIQETTALGAAYLAGLA 449
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 13
; Sequence 12335, Application us/09815242
; Patient No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ. ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 12335
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12335
Query Match
  9.4%: Score 264.5; DB 10; Length 498;
  Best Local Similarity 21.9%: Pred. No. 4.3e-16;
  Matches 112; Conservative 77; Mismatches 227; Indels 95; Gaps 17;

OY 5 KTVIGVDVSGSVRAGIFDLNGLSLSHATEKITTRRSRSGSVEQSSOEIMQAVCSIRNA 64
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 3 KYLISIDOGTSSRAILFNQKEIAGVAQREFKQYFQSGVNEHDANEIMTSVLAVMTEV 62
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 65 LTLADVCAQSVAGIGF-DATCSLVYLDKN-GDPLVSPREGDAKNIITVMDHRTGEQAE 122
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 63 INENDVRADQIAGIGITNORETIVVMDKHGTRPI-----YHAIWQSRQTSICSE 113
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 123 I--NATHNPVLNVYGKISPEMETPKILMLKENNPEIYERAGQ----FFDLADFLTRAT 176
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 114 LKQGYGQTFERDKTGLLDLPYFAGTKVKMLDNVEGAREKANGDILFGTIDTWLVKLS 173
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 177 GDLARSVCTVCKWTWL--AHENRMDPYERTIG----LAELADEDFI--RIGHIIVS 226
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 174 GKAAHITDYSNASTLMFNHIDEMDELLELLTPKRMLEPVKPSSEVYKTIIDYHFG 233
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 227 PGPFCGN-----GLTQAQAEMGILGPFPVAVGLIDAHAGIGITVGEAGALNLAIV 279
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 234 QEVPFIAGVAGDQALRGACFERGDVKNYGTGFMIMNTGD--KAVSESGLTTTAY- 291
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 280 FGTSSCTMASTTSPFVPGVWPYYSAMVPGMLVEGQSAAGAIDQLDFHPAVEAR 339
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 292 -----GIDGKYNVAL-----ESIFVSGSAIQ----- 313
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 340 EMAORVNPPLPWLADRI-LEKTAQPSDAVAL---AKGLHVEPELGNRAPDPHARA 394
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 314 -----WLRDGLRMINSAPQSESYATRVDSIEGVVPAFGLGIPYDSEARG 361
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 395 VICGLGMRDLNLLAIYIAGLCIGYGLRQILDQTA--QGVSKNIYISGAGQHPVLR 453
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 362 AIFGLTGTGTEKEHFIKATLESIC---YQTRDYMAMSKDSIDVQSLRVDGAVKNNFIM 418
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 454 QILADTCGIPYITTCCEPVLLGSAILGAVA 484
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 419 QFOADIYNTSVREPEIQETTALGAAYLAGLA 449
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 14
; Sequence 13788, Application US/09815242
; Patient No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```


Db 437 PVMRETTALGAAYLAGLACGFWSLDE-LKSKAVIERVEPE 477

Search completed: March 13, 2003, 16:51:23
Job time : 17.0298 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:50:47 ; Search time 18.5662 Seconds

(without alignments)
2765.009 Million cell updates/sec

Title: US-09-802-208B-4

Perfect score: 2811

Sequence: 1 MTTKTVIGVDVSGSGSVRAG.....HHRRYEAYKQLQHTAKLLRD 534

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2031	72.3	535	2 S78598	D-ribulokinase (EC
2	1666	59.3	536	2 A99199	D-ribulokinase (EC
3	1666	59.3	536	2 AH3087	ribitol kinase [im
4	1560.5	55.5	538	2 AB3632	D-ribulokinase (EC
5	1244.5	44.3	545	2 AF0442	probable carboxydr
6	1167.5	41.5	569	2 D85354	hypothetical prote
7	971	34.5	525	2 A96183	D-ribulokinase (EC
8	971	34.5	525	2 A13103	ribitol kinase [im
9	864.5	30.8	715	2 S52675	probable membrane
10	465	16.5	509	2 A95985	probable sugar kin
11	448.5	16.0	560	2 D69587	L-ribulokinase ara
12	405.5	14.4	566	2 C90637	L-ribulokinase [im
13	405.5	14.4	566	2 C85488	L-ribulokinase [im
14	402	14.3	563	2 H83883	L-ribulokinase ara
15	394.5	14.0	566	1 K1ECRU	ribulokinase (EC 2
16	381.5	13.6	500	2 D97221	xylose kinase [im
17	377	13.4	569	2 AD0515	L-ribulokinase [im
18	368.5	13.1	530	2 AF0051	probable carboxydr
19	368.5	13.1	569	1 A24984	ribulokinase (EC 2
20	362	12.9	492	2 C72417	sugar kinase, FGGT
21	361	12.8	567	2 AF0274	ribulokinase (EC 2
22	353	12.6	530	2 AE0941	probable sugar kin
23	352.5	12.5	542	2 S63222	MPA3 protein - ye
24	340	12.1	545	2 B89823	hypothetical prote
25	336	12.0	493	2 G69468	carboxydrate kins
26	334.5	11.9	502	2 E83353	xylose kinase Pa
27	328.5	11.7	504	2 AF1789	glucanate kinase h
28	326	11.6	530	2 F90893	probable kinase [l
29	326	11.6	530	2 C85724	probable kinase yd

30	325	11.6	501	1 S18562	xyulokinase (EC 2
31	324	11.5	501	2 G75337	glycerol kinase -
32	323	11.5	530	2 B64905	sugar kinase homol
33	322.5	11.5	499	2 D69735	xyulose kinase xy
34	322.5	11.5	509	2 G95901	probable L-xyulok
35	321.5	11.4	499	2 S74045	glycerol kinase (E
36	320	11.4	504	2 AG1413	glucanate kinase h
37	319.5	11.4	511	2 G98252	cryptic L-xyulose
38	319	11.3	494	2 F83199	probable carboxydr
39	317.5	11.3	524	2 JN0606	ATP-stimulated glu
40	316	11.2	506	2 B72396	sugar kinase, FGGT
41	316	11.2	513	2 D72753	probable xyulose
42	315.5	11.2	492	2 AE3033	L-xyulose kinase
43	314.5	11.2	519	2 E83268	probable carboxydr
44	312.5	11.1	501	2 A99319	transporter/facili
45	310.5	11.0	524	1 S36175	glycerol kinase (E

ALIGNMENTS

RESULT 1
S78598
D-ribulokinase (EC 2.7.1.47) - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 17-Mar-2000
C:Accession: S78598: S08571
R:Heuvel, H.; Shakeri-Garakanl, A.; Turgut, S.; Lengeler, J.W.
Submitted to the EMBL Data Library, January 1998
A:Description: Genes for D-arabinitol and ribitol catabolism in enteric bacteria.
A:Reference number: S78598
A:Accession: S78598
A:Molecule type: DNA
A:Residues: 1-535 <HEU>
A:Cross-references: EMBL:AF045244
A:Experimental source: Strain KAY2026
R:Loviny, T.; Norton, P.M.; Hartley, B.S.
Biochem. J. 230, 579-585, 1985
A:Title: Ribitol dehydrogenase of Klebsiella aerogenes. Sequence of the structural ge
A:Reference number: S07135; MID:86050424; PMID:2933028
A:Accession: S08571
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-16 <LOV>
A:Experimental source: strain FG5
C:Genetics:
A:Gene: rbtK
C:Function:
A:Description: catalyzes phosphorylation of D-ribulose to D-ribulose-5-phosphate
C:Superfamily: ribulokinase
C:Keywords: phosphotransferase

Query Match 72.3%; Score 2031; DB 2; Length 535;
Best Local Similarity 72.3%; Pred. No. 2.1e-144;
Matches 384; Conservative 49; Mismatches 98; Indels 0; Gaps 0;

QY	4	TKTVIGVDVSGSVRAGIIPDLNGLSHATEKITTTTRSGSRVDSOELIOWAVCSIRN	63
DB	5	TQNTIGVDVSGSVRAGVFNLRGELLAHATREITLFSSAGNKVQSSREITWQAVCYCIKT	64
QY	64	ALTIADVCAQSVAGICGFATCSLVLDKNGDPLVPSPEGDAKONIIYMDHRRATGEQARI	123
DB	65	AVANAGVSPSSIGIGFATCSLVVIGDNDAPLAVGFSDDADRNIITVMDHRRATGEQARI	124
QY	124	NATHHPVLNVGKISPEMETPKILWLEKMPETIYERAGOFFDLADFLYWRATGDLARSV	183
DB	125	NATGHVPLRVVGVKISPEMETPKILWLEKMPETIYERAGOFFDLADFLYWRATGDLARSV	184
QY	184	CYTTCCKWTVAHNRMPDVFRTITGLADEDFIRGHIITVSGTCGNGGLTAQAAEM	243
DB	185	CYTTCCKWTVAHNRMPDVFRTITGLADEDFIRGHIITVSGTCGNGGLTAQAAEM	244
QY	244	GLLPGRFVAVGLIDAHAGIGITGVVEGALNMLAYVGTSSCMASSTSPSPVGVWGPY	303

```
Db 245 GLPIGTAAAGMIDAHAGGIGTGTGVLGAVNNMAYVFGTSSCTMTTQOEAFFVGVWGPY 304
QY 304 YSAAVPEGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNPOLPWLADRIEKTAAQ 363
Db 305 YSAAVPEGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNPOLPWLADRIEKTAAQ 364
QY 364 PSDAVALAKGLHVVPEFLGNRAPRAPPARAVITGIGMERDLDNLALYTAGLGGIGYGL 423
Db 365 PSEAVTLAAGLHVVPEFLGNRAPRAPPARAVITGIGMERDLDNLALYTAGLGGIGYGL 424
QY 424 RQILDAGTAQGVSKNIVISGAGQHPVLVQILDTCGIPVITTCOCEPVLLGSAIIGAV 483
Db 425 RQILDAGTAQGVSKNIVISGAGQHPVLVQILDTCGIPVITTCOCEPVLLGSAIIGAV 484
QY 484 AGNIAPSVGEAMQOFTVHDKYVYQOERQSLHRRYEAYKOLQHTAKLRD 534
Db 485 AGRVAALPEAMKQFTQVDATYHSETAFSPULHRRYAAKALQOAGRLIRE 535
```

RESULT 2

A99199
D:ribulokinase (EC 2.7.1.47) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: A99199
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: A99199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89115.1; PID:g15158921; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1075
A:Map position: linear chromosome
C:Superfamily: ribulokinase
C:Keywords: phosphotransferase

Query Match 59.3% Score 1666; DB 2; Length 536;
Best Local Similarity 60.0%; Pred. No. 4.9e-117; Mismatches 146; Indels 0; Gaps 0;
Matches 316; Conservative 65; Mismatches 146; Indels 0; Gaps 0;
QY 7 VIGDVGSGSVRAGIFDLNGLSLSHATEKITTTTTRSGRVSQSOEIQAVCSIRNALT 66
Db 5 LVGVDTGTSARAGVFDVAGKLTAKRPISMHREDGIAQSSAEVQAVCDSDVS 64
QY 67 LADYCAQSVAGIGDATCSLVYLDKNDPLVPSPEGAKONIYWMHRAVEQAEIRINAT 126
Db 65 RAGIDPAEVTIGIGDATCSLVYRGRGVDVTLVPGAADHPERDIYWMHRAVEQAEIRINAG 124
QY 127 HHPLVNVGKISPEMETPKILMKENMPETERYRAGOFDLADFLTPMATDGLARSCTV 186
Db 125 KHAVLKTYGGRISPEMOTPKILMLSENRPDIYAAAEHFDLTDPLTKKASALDRSACTV 184
QY 187 TCKMTWLAHENRMDPDYFRITGLAELEDEDFIRIGHIIVSPGTCGNGLTAAQAAEMGL 246
Db 185 TCKMTWLAHENRMDAEYFTKIGLDLAEQGRIRIGESVYHVGTAALNGLTBEAKAKGLV 244
QY 247 PGTIVAVGLIDAHAGGIGTVEGALNNLALYVGTSSCTMASTTSFVGVGWGPYISA 306
Db 245 AGTAAVAGLLIDAHAGGIGTVAAGDASRCGLYVEGTSSCTMTTTEPAFVGVGWGPYISA 304
QY 307 MVRGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNPOLPWLADRIEKTAAQPSD 366
Db 305 MVRGAMLVNEGQSAAGAIDLDLDFHPAVEAREMAQVNPOLPWLADRIEKTAAQPSD 364
QY 367 AVAALAKGLHVVPEFLGNRAPRAPPARAVITGIGMERDLDNLALYTAGLGGIGYGLRQI 426
Db 365 AVKLAEDFHNVPFLGNRAPRAPPARAVITGIGMERDLDNLALYTAGLGGIGYGLRQI 424

```
QY 427 LDAQTAQGVSKNIVISGAGQHPVLVQILDTCGIPVITTCOCEPVLLGSAIIGAV 486
Db 425 IETQARNGAPVETISVSGAGAHPLARQLADATGRLPELTCEPEVLLGSAMIGAVAG 484
QY 487 IAPSVGEAMQOFTVHDKYVYQOERQSLHRRYEAYKOLQHTAKLR 533
Db 485 TYRPLMAAMPMSRIDSSAMPDPDFOKIHOARYDAFLALQAAARAIR 531
```

RESULT 3

AH3087
ribitol kinase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH3087
R:Mood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woeherge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayvin, T.; Levy, R.; Li, M.; McClintock, P.; Kometani, P.; Zhang, S.; Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH3087
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <KUR>
A:Cross-references: GB:AE00688; PIDN:AL45118.1; PID:g17742788; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4324
A:Map position: linear chromosome
C:Superfamily: ribulokinase

Query Match 59.3% Score 1666; DB 2; Length 536;
Best Local Similarity 60.0%; Pred. No. 4.9e-117; Mismatches 146; Indels 0; Gaps 0;
Matches 316; Conservative 65; Mismatches 146; Indels 0; Gaps 0;
QY 7 VIGDVGSGSVRAGIFDLNGLSLSHATEKITTTTTRSGRVSQSOEIQAVCSIRNALT 66
Db 5 LVGVDTGTSARAGVFDVAGKLTAKRPISMHREDGIAQSSAEVQAVCDSDVS 64
QY 67 LADYCAQSVAGIGDATCSLVYLDKNDPLVPSPEGAKONIYWMHRAVEQAEIRINAT 126
Db 65 RAGIDPAEVTIGIGDATCSLVYRGRGVDVTLVPGAADHPERDIYWMHRAVEQAEIRINAG 124
QY 127 HHPLVNVGKISPEMETPKILMKENMPETERYRAGOFDLADFLTPMATDGLARSCTV 186
Db 125 KHAVLKTYGGRISPEMOTPKILMLSENRPDIYAAAEHFDLTDPLTKKASALDRSACTV 184
QY 187 TCKMTWLAHENRMDPDYFRITGLAELEDEDFIRIGHIIVSPGTCGNGLTAAQAAEMGL 246
Db 185 TCKMTWLAHENRMDAEYFTKIGLDLAEQGRIRIGESVYHVGTAALNGLTBEAKAKGLV 244
QY 247 PGTIVAVGLIDAHAGGIGTVEGALNNLALYVGTSSCTMASTTSFVGVGWGPYISA 306
Db 245 AGTAAVAGLLIDAHAGGIGTVAAGDASRCGLYVEGTSSCTMTTTEPAFVGVGWGPYISA 304
QY 307 MVRGLMVEGQSAAGAIDLDLDFHPAVEAREMAQVNPOLPWLADRIEKTAAQPSD 366
Db 305 MVRGAMLVNEGQSAAGAIDLDLDFHPAVEAREMAQVNPOLPWLADRIEKTAAQPSD 364
QY 427 LDAQTAQGVSKNIVISGAGQHPVLVQILDTCGIPVITTCOCEPVLLGSAIIGAV 486
Db 425 IETQARNGAPVETISVSGAGAHPLARQLADATGRLPELTCEPEVLLGSAMIGAVAG 484
QY 487 IAPSVGEAMQOFTVHDKYVYQOERQSLHRRYEAYKOLQHTAKLR 533

Db 485 'TYPDIMAAMPANRIDSAMPDPDFOKIHOARYDAFLALONARAIR 531

RESULT 4

AB3632

D:ribulokinase (EC 2.7.1.47) [Imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 17-May-2002

C:Accession: AB3632

R:DelVecchio, V.G.; Kapural, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AB3632

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-538 <KUR>

A:Cross-References: GB:AB008918; PIDN:AAL54221.1; PID:g17985192; GSPDB:GN00191

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME10979

A:Map position: 11

C:Superfamily: ribulokinase

C:Keywords: phosphotransferase

Query Match 55.5%; Score 1560.5; DB 2; Length 538;

Best Local Similarity 57.8%; Pred. No. 4e-109;

Matches 309; Conservative 66; Mismatches 155; Indels 5; Gaps 3;

2 TTTKTVIGVDSGVRAGIFDLNGLSLSHATEKITTTTFRSGSRVEOSSQEIWOAVCSCT 61

4 TMTNYLVGVDVCTGSARAGLFDAGGTMLASARDIAIWMRAGIVGSSDDIMQANCEV 63

62 RNALVLADYCAOSVAGIGFDATCSLVLDKNDPLVSPBGAKONIYWMHRAEQAE 121

64 REVNVAGVDPRAVAGIGFDATCSLVLDKNDPLVSPBGAKONIYWMHRAEQAE 123

122 RNTAHHPPLVNVGKISPEMETPKILMKENMPETIYERAGOFPLDPLTRATGDLAR 181

124 RINTTKADVLGVGALISEMETPKILMKENMPETIYERAGOFPLDPLTRATGDLAR 183

182 SVCTVTKMTWLAHNRMPDYFRTIGLAELADEDFIRIGHIIVSPGTCNGITQAQAA 241

184 SACTVTKMTWLAHNRMPDYFRTIGLAELADEDFIRIGHIIVSPGTCNGITQAQAA 242

242 EMGLLPRTAVAGLIDAHAGIGTVGV--EGALNMLAVFEGTSSCTMASTTSPSEVPG 298

243 ELGLRPGTAIAAGLIDAHAGIGTVGV--EGALNMLAVFEGTSSCTMASTTSPSEVPG 302

299 VNGPYTSANVPGLMLVEGOSAGAALDQLDFHRAVEAREMAQRVNPPLVWLADRL 358

303 VNGPYTSANVPGLMLVEGOSAGAALDQLDFHRAVEAREMAQRVNPPLVWLADRL 361

359 ERTAOPSDAVNALAKGLHNVPEFLGNRAPADPHARAVICGLMERDNLMLALYIAGLCG 418

362 EARGPEPKTAMIVGDIHVPEFLGNRAPADPHARAVICGLMERDNLMLALYIAGLCG 421

419 IGYGLRQILDAGTAAOVVSKNIVISGAGOHPLVROLADTGCIPVITTOCCPEVLGSA 478

422 LGYGLRQILDAGTAAOVVSKNIVISGAGOHPLVROLADTGCIPVITTOCCPEVLGSA 481

479 ILGAVAGNIAPSVGEAMQOFTHVDKYVYQERVQSLHHRVYAYKQLOHTAKILR 533

482 MIGAVALSGAYPDLVYAMVMSSELGARNRPDARRAKMWHDRFAEMLLQATARKIR 536

RESULT 5

AF0442

Probable carbohydrate kinase YPO3637 [Imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AF0442

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.

demo-Tarrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AF0442

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-545 <KUR>

A:Cross-References: GB:AL590842; PIDN:CAC93106.1; PID:g1598158; GSPDB:GN00175

C:Genetics:

A:Gene: YPO3637

Query Match 44.3%; Score 1244.5; DB 2; Length 545;

Best Local Similarity 45.3%; Pred. No. 2e-85;

Matches 241; Conservative 100; Mismatches 174; Indels 17; Gaps 5;

8 IGVDSGVRAGIFDLNGLSLSHATEKITTTTFRSGSRVEOSSQEIWOAVCSCTRNALTL 67

6 IGVDSGVRAGIFDLNGLSLSHATEKITTTTFRSGSRVEOSSQEIWOAVCSCTRNALTL 65

68 ADVCAOSVAGIGFDATCSLVLDKNDPLVSPBGAKONIYWMHRAEQAEARINATH 127

66 ADINPIQVKGIGFDATCSLVLDKNDPLVSPBGAKONIYWMHRAEQAEARINATH 125

128 HPLVNVGKISPEMETPKILMKENMPETIYERAGOFPLDPLTRATGDLARSCVTV 187

126 HPLVNVGKISPEMETPKILMKENMPETIYERAGOFPLDPLTRATGDLARSCVTV 185

188 CKMTWLAHNRMPDYFRTIGLAELADEDFIRIGHIIVSPGTCNGITQAQAAEGLLP 247

186 CKMTWLAHNRMPDYFRTIGLAELADEDFIRIGHIIVSPGTCNGITQAQAAEGLLP 245

248 GTPVAVGLIDAHAGIGTVGV--EGALNMLAVFEGTSSCTMASTTSPSEVPG 301

246 GTPVAVGLIDAHAGIGTVGV--EGALNMLAVFEGTSSCTMASTTSPSEVPG 305

302 PYTSANVPGLMLVEGOSAGAALDQLDFHRAVEAREMAQRVNPPLVWLADRL 357

306 PYTSANVPGLMLVEGOSAGAALDQLDFHRAVEAREMAQRVNPPLVWLADRL 360

358 LEKTA-OPSDAVALAKGLHNVPEFLGNRAPADPHARAVICGLMERDNLMLALYIAGL 416

361 LQMGAPENIAFLTDIMHLPFHGNSRBRANPLTGITGLKSTTEBDMLARLATI 420

417 CGIGYGLRQILDAGTAAOVVSKNIVISGAGOHPLVROLADTGCIPVITTOCCPEVLG 476

421 GMLALGTRHILIEFMNONGNIDTMMASGGCTKPIVYOEHAANTGCAMLLPESEAMLLG 480

477 SALGAVAGNIAPSVGEAMQOFTHVDKYVYQERVQSLHHRVYAYKQLOHTAKILR 527

481 SAMGTVAGVPESTLEPAAAMASRIGKTVTPQTNKIKAYDRKRVVHOMYH 532

RESULT 6

DB5354

Hypothetical protein AT430310 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: DB5354

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: AB5001; MUID:20083488; PMID:10617198

A:Accession: DB5354

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-569 <STO>

A:Cross-References: GB:NC_001268; MUID:g7269931; PIDN:CA81024.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT430310

A:Map position: 4

Query Match similarity 41.5%; Score 1167.5; DB 2; Length 569;
Best Local Similarity 44.0%; Pred. No. 1.3e-79;
Matches 242; Conservative 95; Mismatches 178; Indels 35; Gaps 9;

OY IGVDSGGSVRAGIFDLNGLSLSHATEKITTTTRRSRVEOSSOEIMOVACSICINNALTL 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB LGVDVGTTSARAGIEDDDNGKLLGSSATSPFIQW-KGDGCIEGSTFIDMVAHCAAVSACL 74
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY ADVCGASVAGTGFDAITCSLVYLDKRGDPLPVSPEBDAQIONITYMMDHRTBDAAEILNTH 127
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB ANMSDVYKGGIGFAATCTSLVAADAGESPVTWSGSDSRNNITVMMDHNAVKOAERINSFN 134
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY HPVLNVYSGKISPSMETPKILLKENMPEIYERACQFDLADFELTPRAQTGLDARSVCYT 187
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB SPVLDTGCGGCVSPENEPKLLMKVENLKESMSMYKKMNDLSWLSTYRATGDOTRSLCTTV 194
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY CKWTMLAHENR-----WDPDYFRITGLAELEADEDFIRIGHIVSFPTPGC 232
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB CKWTGLGHAMHMOMTEKASRDMEACGWDEFEWEELIGLDLYDGHNHAKIGRSVAFFGNPLG 254
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY NGLTQAQAEMGLPGRPAVAGLIDAHHGISTVGEAGALNL-----AYVGSTSCMA 288
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB NGLTITTAKEIKELLGTPVGTSLIDAHAGVGY--MEKSDVDTLCSRMLVCGTGTCMA 312
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY STTSPFPVPQWGPYYSAMVPGLMLVEGQSAGAAILDIQLDFHPAVEAREMARVNOP 348
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB VSREKFLIPQWGPFWMSAMVPEWTLETGGQSATGALLDHIIENH--VASPRIANNAASOK 370
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY LPYW-LADRIELEKTQPDSA---VALAKGLHVREFLGKNRAFDPARNARAIVTGLSMERD 404
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB VSVFELLNNTIKTMAEDTSSPFISALTSDMHILPDFHGRRSPVADPNNSKVTFEGMSLPTS 430
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY LDNLATLYTAGICGIGYGROIIDAQTAGQVYSKNIVISGAGOHPIRQILDADCPGV 464
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB EKQALLTLATIQGTAGTGRHTVEHCNTHGKIIDLKACGGLSKNPRLFIDEHADIVGPT 490
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY ITTQCCEPVLLSALIVAGNAPSVGEAMOQFTNHVDKYXXYPQE--RYOSLHHRRYEAY 522
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB ILPRESESVLLGAAILGAVAGNKYPSLDAMKALMAAGVYNRPSSDPRIKKKHDAKKYRF 550
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 523 K-----QLOH 527
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 551 RNLYEQQLSH 560
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 7
A96183
Description: Dribucklinase (EC 2.7.1.47) [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
Species: Agrobacterium tumefaciens
Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
Accession: A96183
R:Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Marketz, B.
Science 294, 2323-2328, 2001
Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens 294, 2323-2328, 2001
Reference number: A97359; PMID:11743194
Accession: A96183
Status: preliminary
Molecule type: DNA
Residues: 1-525 (KOR)
Cross-references: GB:AE007870; PIDN:AAK88987.1; PTD:g15158771; GSPDB:GN00170
Genetics:
Gene: AGR_L_826
Map position: linear chromosome
Keywords: phosphotransferase

Query Match	34.58%	Score 971;	DB 2:	Length 525;
Best Local Similarity	40.08%	Pred. No.	6.1e-65;	
Matches 216;	Conservative 91;	Mismatches 187;	Indels 46;	Gaps
5 KTVICGVGSGSRAPIFDNGLSLSHAETKTTTRRBSGRVSSQEIQAVSCIRNA 6	:	:	:	:
: :: : :::	:	:	:	:
:	:	:	:	:

Db	4	QNLVAVDVSTASARGIEFDTPAGRIILARSINPLIOMQRPREHNAEDHSDINDMAVCSIAKAA	63
Oy	65	LTLDVQCSQSVAGIGFDRATCSLVYLCKNGDRPLPVSREDDAQONITIMMDHATQOABERIN	124
Db	64	LADAVLEPQSTIAAGIEDTCSLVIRDERGEVSVSTDDDDFTVITLNDHNAIEADPLT	123
Oy	125	ATIHNEVLNVQSGKISPEAREPKILMLKEMPEIYERAGQFRLDPLFWIRATGDLARSVC	184
Db	124	ASGHRVLEDFAGNSVSRPEQMKIKMMLKTHMPOQSMRMSFARDLADFLTWKATGSAQNSNC	183
Oy	185	TVTCKMTVLAHEN-RMDDPYFRITIGLAELADEDPTRIG--NHIVSPGPGSNGITLQAAA	241
Db	184	TOTALMNLAEHNPQWQADYLAFAGLDPLKE---RAGLPETVWPGGALG-PLSPREAAA	238
Oy	242	EMGLRPTGVAVUGLIDAHAGIGITVGVGEAGL-----NNLAVUEFTSCTMASTTSPSFV	296
Db	239	ELGIDTGGQVAAQMDIAYAGALGAL--GGLAEUDCKHVALIGTSGSLVAMSTOP--M	293
Oy	297	PG--VMGRTYSAMVGTMLVVEGGSAGAGALDOLLDFPRAVEAREMAQKNOPLRWMLA	354
Db	294	PGRSITMGRYWQAVLEGHMLVEGGSATGALLDHLIVRMHAAGER-----DTALHARIV	346
Oy	355	DRLEKTPQSPDAVVLAKGLHVHVEPEFLGNRAPFADPHNAAYICGLGMRDLDNLATLUTYA	414
Db	347	ARVTE--LRELEGEFARLRLHVLPDFHGNRSRLADPRNAVGVSGTLDTSPDISCLRUYMR	404
Oy	415	GLCGIGYCLROLDLQATQGVVSKNIVISGAGOHRLVROLADTCGIVITTCOCSEVYL	474
Db	405	TAIATAIARHVLUDMEFEYGAUETLHVTCGHVKNRPLMELVADVTGKRIYVPPATAVVL	464
Oy	475	LSGATLGVAVNGIAPSV---GEAMQO-----FTYHDKYUYUPOERYOSLHNRRE	520
Db	465	LGTANTATAGGVAHSLAAGAGAMYPGNAETSGNPALAHNERDY---RRLAMYKRRQOE	521

```

ribitolc_kinas[imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: A13103
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Mons, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guentherer, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: A13103
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-525 <KSR>
A:Cross-references: GB:AE008689; PIDN:AAL45247.1; PID:g1774231; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4453
A:Map position: linear chromosome

```

	Query Match	34.5%	Score 971;	DB 2;	Length 565;	
	Best Local Similarity	40.0%;	Pred. No. 6.1e-65;			
	Matches 216; Conservative	91;	Mismatches 187;	Indels 46;	Gaps 13;	
OY	5 KTVIGVDVGGSSVRAGIFDNLGSLLSHATEKITTTTTRRSGSRVEQSQELIOWAVCSIRNA	64				
	: : : : : : : :					
Dd	4 ONLVADVDTASARAGIEDDPAGRLRLARSIIHPILMORPRENHAEHSDTDIMNAVCIAYKAA	63				
OY	65 LTLADVCNOSVAGIEFDATCSLVLADKNDDPLPVSPEGAKNIITYMDHRATEQAERIN	124				
Dd	64 LADAGVLPGQSIAAIEFIDATCSLYRIDEREPEVSVSTTDDRDREDTIWMLDHRAIGEARDLT	123				
OY	125 ATTHGVLLNVVGSKISPEMTPKIIMLKEMMPITERYACGFEDLADFLTYRAAGDLARSC	184				
	: : : : : : :					

Db	124	ASHRVLDPAGNSVSPBEMQPKLMLKLNHPQSMKMSKSPADLPDLFLTKAKGSAQNRNC	183
Oy	185	TYTKCKTWLAHEN-RWDPDYFRTIGLAELADEDFIRIG--NHIVSPGTPCGNGLTQAQAA	241
Db	184	TOTAKNPLAEQHEHQMOADYLAFAGLDDLKE-----RAGLPETTPMGGAIG-PLSPFEEAA	238
Oy	242	EMGLLPSTPAVGLLIANHAGIGSTVVEGAL-----NNLAVFEGISSTMASTSPSFV	296
Db	239	ELGLDTCQYAAAGNIDYAAAGALAL--GSGCLAEVKGKIVALIAGTSSCLVMASTQP--M	293
Oy	297	PG--VMGPRYSAMYPGLMYLVEGGOSAAQAAIDOLDPFRHVEAREBMAQRYNPRLWILA	354
Db	294	PKRSLMKPRYQAVLCPRGMYLVEGGQSTGALLDHYVMHNAAGSR-----DTLHNAIV	346
Oy	355	DRILEKTAQPSDAVALAKLGHVVPPEFLGNKRAPFRADPHARAVYICGLGEMERDLDNLATYA	414
Db	347	ARVTE--LRELEGBAFADRLHVLPRDHGNKNSPLADPHNAVGVSGLLDTSFDSLRLWMR	404
Oy	415	GLCGTGTGLKQIILDAQTAGVSSKNIVISGAGQNRVLKQIILADPGIRPVITTTQSCSEVYL	474
Db	405	TATAIAGAARHVLDAMERFQYAVETLHVTVGHVKNPRLLELADYVTKRIIVPATADAVL	464
Oy	475	IGSAIILGAVNGLAPSV-----GEAMQO-----FTNVQKYUUPQEROQSLNNHRYE	520
Db	465	LGTMATATATGCVNASTAAAGAAATGCAENAEISGNPLAAHNERDY---RRFLAMTRHNOE	521

RESULT 9

Probable membrane protein YDR109c - Yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein YD9727.05c
C:Species: *Saccharomyces cerevisiae*
C:Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S52675
R:Murphy, L.; Shore, L.; Harris, D.
submitted to the EMBL Data Library, March 1995
A:Reference number: S52671
A:Accession: S52675
A:Molecule type: DNA
A:Residues: 1-715 <MUR>
A:Cross-references: EMBL:Z48758; NID:g747879; PID:g747884; GSPDB:GN00004; MIPS:YDR109c
C:Genetics:
A:Gene: MIPS:YDR109c
A:Cross-references: SGD:S0002516
A:Map position: 4R
C:Keywords: transmembrane protein
E:515-531/Domain: Transmembrane #status predicted <TM1>
E:535-551/Domain: Transmembrane #status predicted <TM2>

[illegible]

Qy	274	NNLIVFSSCTMASTSPFSFVGVWGPYYSAMVPGILMVEGGGSAGAAIDOLDJHP	333
Dd	337	GRLAVACSTICHILLSNPIFVHCWGMPRYDLARGMAABEGGSCGLVDHLITTHP	396
Qy	334	AVEAREMAORVNOPLPYMLADRLIEKTAQP--SDAVALAKGHVPEFLGNRAFEADP	390
Dd	397	AFTLSHMANIAGVSKEFYEL-NKILETLEVKRKVRYSISIAKLHLPFYGDYGHSNPSIIDP	455
Qy	391	HARAVICGLMERDLDNLALYLINGLCIGTGRLRIIDAQAOGVSKNIYISGAGGHP	450
Dd	456	NMRACIIIOOSMDNSTIEDLAVMYLSACEPTISOOTROIIEVMUKSGHEINAIATMSGGOCNS	515
Qy	451	LVRQIIDLTCGIPLYTTQOCE-PVLLGSAILGAVANAPASVGEMAOQFTHVDKYYXPQE	509
Dd	516	LLMRLDACGTGPIYIPRYVDAVAVFSGALLGAASAEDF-----DYTBREKRLLKGXK	567
Qy	510	RYSOLHHRRRYEAYKQLQ	526
Dd	568	SSQYTERFNDSYSIQ	584

RESULT 10

Probable sugar-kinase protein (EC 2.7.1.-) [imported] - *Sinorhizobium meliloti* (strain C) Date: 24-Aug-2001 #sequence,revision 24-Aug-2001 #text,change 30-Sep-2001
C:Accession: A95985
C:Rifan, T.M.; Weldon, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her-
proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSYM6 megaplasmid from the N2-fixing e-
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A95985
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-509 <KDR>
A:Cross-references: GB:AL591985; PIDN:CAC49545.1; PID:g15141032; GSPDB:GN00167
A:Experimental source: GB:AL591985, strain 1021, megaplasmid pSYM6
R:Galibert, F.; Rifan, T.M.; Long, S.R.; Punher, A.; Abola, P.; Ampe, F.; Barloy-Hubbl-
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau-
nehalt, P.; Vandenbol, M.; Vorholter, F.J.; Weldon, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SM20852
A:Genome: plasmid
A:Keywords: phosphotransferase

```

Query Match 16.5%; Score 465; DB 2; Length 509;
Best Local Similarity 26.3%; Pred. No.5.2e-27;
Matches 133; Conservative 89; Mismatches 213; Indels 70; Gaps 16;

OY 7 VIGVUDGSGSVACGIFDL-NGSLSHATEKITTTRRSGRSVDSOSELQWACSCIRNAL 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 4 VLSDIFGIGGARAALIFDQTNINIVGAEARKYGTQNHLPRAQDNEDMMTALVSLVRVY 63
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 66 TLA---DYCAQSVAGIGFDTATSLVVLDKNCPLRVSPRGDAKONTIYMDHRAIEQAR 122
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 64 AKAGSPDIAACVAF---ASTVVLCDRSGK-PIAP-----AVLMDRADDEAARF 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 123 INATHNRLVNYGKGISPEMERPKILMLKENNREIYERAGOFPLDLPTWATGTLARS 182
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 111 TETVDHPLTADSGSDAAEWMLVLRKMMWRARRKRDIMARTEVICALDPLNNHRTGWAOS 170
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 183 VCTVYCKTWTLANENRMDPRDYRTIGLALA---DEDIFRIGHNIYVSPGCGNGLTAOA 239
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 171 LMNATTCCKNNYDRNRKRFCDLIALRGVPLDGAKLPRQRIYDG-DVAVRPLP-----EM 222
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 240 AAEMGLKPTPLVAV-GLIDAHAGIGIVGVEGALNNLNAVPGTSSCTMASTTSPS-FVP 297
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```


Db 336 SAEGDIYAMGRLVGLPMLQALAEHLKQIDASOK--QLPALTEAMAKNSLDH--- 390
QY 364 PSAAVALAGLHLYPPE-FLGNRAPFPADPHARAVYICGLGMRDLNLLALYIAGLCIGYG 422
Db 391 -----LPVLLDMFNGRRTPMANQRLKGYITDNLATDAPLLFGGLIA---ATAFG 437
QY 423 LRQILDQAOTAGVSKNIVISGG-AGOHPLVRQILDATGCP---VITTOCCPEVLLGSA 478
Db 438 ARAIMCEFTDQGIANNVNMALGGIARKNOVIYMACCDVLRPLQIYASDQCC---ALGAA 494
QY 479 ILGAVAGNIAPSVGEAMQOF-THVDKYYYPQ---ERYOSLHRRREAY 522
Db 495 IFAVAIAKVAHADIPSAQOKMASAVEKTLQPRSDQAQREQL-VRRYQOW 542

RESULT 13

C85488
L-ribulokinase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85488
Riperna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimlantia, E.; Potamoculis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85488
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <STO>
A:Cross-references: GB:AE005174; NID:g12512760; PIDN:ACG54367.1; GSPDB:GNO0145; UMGPR:200
C:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: arab
C:Superfamily: ribulokinase

Query Match 14.4%; Score 405.5; DB 2; Length 566;
Best Local Similarity 27.7%; Pred. No. 1.7e-22;
Matches 163; Conservative 77; Mismatches 224; Indels 125; Gaps 29;

QY 8 IGVDVSGSVYRAGIFDLNGLSLSHATEKITTRRSGRVDSOEOIWO--AVCSIRNAL 65
Db 5 IGLDFGSDSVYRALVD-----CTTGELIATSVEMYP--WQKQCFDAPANNQF 50
QY 66 -----TLADYCAQ---SYAGIGFDPATCL-VLIDKNGRPLVSPGDA 104
Db 51 RHHRPDYIESMEALKTIVLAELSEVEDRAAVVIGVDSGTSTPAPIDADGNVLAIRPEFAE 110
QY 105 KON-IIVMDHRAEQERINATNHPVLN-----YVGGKISPEMETPKILMLKENPEI 157
Db 111 NPMAMVLMKDHRAVEAEETILCHAPGNVDYSRVIGIYSSEWFAKILHTRQDSAV 170
QY 158 YERAGOFDLADFLTWBRATG-----DLARSVCVTCTKWTWLAHNRMDPYFTTGL--A 210
Db 171 AOSAAISMIEICDVPVALLSGTTRPDIRGRCSAGHKSIM--HES-WG-----GLPPA 220
QY 211 ELADEDFIRGHHIVSP-----GTPCGNGLTQAQAAAMGILLGPRFVAVGLIDAHAG 262
Db 221 SPFDELDPLNRLHPLSLFTDTYADIPVGT-LCPENAKRLGIPESVVISGGAFCDHGA 279
QY 263 IGVVEGEGALNMLAVVFTSCTMASTTSPSF---VPGVMPYYSAMVPGMLVVEGQ 318
Db 280 VCA-----GAQPNALVKVIGISTCDILADKOSYGERAVKVICQGVDSVYVPGFGLGAGQ 335
QY 319 SA-----AGAAIDQLDFHRAVEAREMAQRVNPPL---VWLADRILEKTAQ 363
Db 336 SAEGDIYAMGRLVGLPMLQALAEHLKQIDASOK--QLPALTEAMAKNSLDH--- 390
QY 364 PSAAVALAGLHLYPPE-FLGNRAPFPADPHARAVYICGLGMRDLNLLALYIAGLCIGYG 422
Db 391 -----LPVLLDMFNGRRTPMANQRLKGYITDNLATDAPLLFGGLIA---ATAFG 437

QY 423 LRQILDQAOTAGVSKNIVISGG-AGOHPLVRQILDATGCP---VITTOCCPEVLLGSA 478
Db 438 ARAIMCEFTDQGIANNVNMALGGIARKNOVIYMACCDVLRPLQIYASDQCC---ALGAA 494
QY 479 ILGAVAGNIAPSVGEAMQOF-THVDKYYYPQ---ERYOSLHRRREAY 522
Db 495 IFAVAIAKVAHADIPSAQOKMASAVEKTLQPRSDQAQREQL-VRRYQOW 542

RESULT 14

H83883
L-ribulokinase arab [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: H83883
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; I
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans e
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: H83883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-563 <STO>
A:Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05591.1; GSPDB:C
A:Experimental source: strain C-125
C:Genetics:
A:Gene: arab
C:Superfamily: ribulokinase

Query Match 14.3%; Score 402; DB 2; Length 563;
Best Local Similarity 26.7%; Pred. No. 3.1e-22;
Matches 155; Conservative 82; Mismatches 260; Indels 84; Gaps 20;

QY 4 TKVIVGVDSVYRAGIFDL-NG-SLSHATE-----KITTRRSGRVDSOQ 51
Db 3 TKYITGVDTGEGRAVLIDLNSQELADHVTPRHGVIDOYLPTNTIKLGHEMALQHP 62
QY 52 EIMQAVCSIRNALTLADYCAQSVAGIGFDPAT-CSLVVLDKNGDPLVSPGDAKON--- 107
Db 63 DYVEVLITSPRAYMKESGVADVDYIGVDPFTACTMLPVDEGQPLCLLAQ--YKDPHS 120
QY 108 -IIVMDHRAEQERIN---ATNHPVLNVGKISPEMETPKILMLKENPEIYERAG 162
Db 121 VWKLMKHNHAQODKANAINEAKKEAFLEPRYGKISSEWMAKVMQILDEADVYNRD 180
QY 163 QFEDLADFLTWKATGDLARSVCYTCTKWTWLAHNRMDPYFTTGLAELEDFIRIGH 222
Db 181 QFLEATDVIYSOMGKIVKNSCTAGYKAIWKKRGYPSNEFYALD--PRLHLTTTKLRG 239
QY 223 HIVSPGTPCGNGLTQAQAAEMGILLPTPVAVVGLIDAHAG---GIGTVEGEGALNMLAY 278
Db 240 DIYPLGSRAG-GLPENAEMKGLNPGIAVAVGANDAAAPAVGVTTTPG-----KLVM 291
QY 279 VFGTSSCTMASTTSPSFVPGVMPYYSAMVPGMLVVEGQSAAGALIDQLDHPRAV--- 335
Db 292 AMGSTISMLLGEKEDEVEEGMCGVEEDGIIIPGYLGAGQSAVGDIPAMVVKHGSATF 351
QY 336 EEARERAMQRNPLRPVLAIRILEKTAQ--PSDAVALAKGLHLYPPEFLGNRAPADPHAR 393
Db 352 DEAOEKGVNVALE-----EKASOLRPGES-----GLAIDMWNNGRSLIVDELS 398
QY 394 AVICGLGMRDLNLLALYIAGLCIGYGLRQILDQAOTAGVSKNIVISGGAQO--HPLV 452
Db 399 GMLGTYLQTKRPE---IYALLETATFGRATRYADAHNGRGVNEHLXACGGRLQKNNHL 455
QY 453 RQILDATGCPVITTOCCPEVLLGSAIIGAVAGNIA---PSVGEAMQOF----- 498
Db 456 MQIPADVTNREIKVAASKQRPALGAAMFASVAAQSEVGYDSIREAKKMKRWYDETFKP 515
QY 499 --THVDKYYYPQERYOSLH-----RRTAYVQOLH 527
Db 516 IPEHVAIYERLYQERYVTLHDYFGRGANDVMKRLKALKSIQH 556

RESULT 15
KIECRU
ribulokinase (EC 2.7.1.16) - Escherichia coli (strain K-12)
N:Alternate names: L-ribulokinase
C:Species: Escherichia coli
C:Date: 30-Jun-1968 #sequence-revision 31-Oct-1997 #text-change 01-Mar-2002
C:Accession: G64727; B29022; S40579; I41134
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
A.: Rose, D.J.; Mau, B.; Shao, Y.
S:Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64727
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-566 <BLAT>
A:Cross-references: GB:AE000116; GB:U00096; NID:g1786240; PIDN:AAC73174.1; PID:g1786249;
A:Experimental source: strain K-12, Substrain MG1655
R:Lee, N.; Gielow, W.; Martin, R.; Hamilton, E.; Fowler, A.
Gene 47, 231-244, 1986
A:Title: The organization of the arabid operon of Escherichia coli.
A:Reference number: A91559; MUID:87163495; PMID:3549454
A:Accession: B29022
A:Molecule type: DNA
A:Residues: 1-22, 'S', 24-126, 'RS', 129-349, 'S', 351-364, 'A', 366-402, 'S', 404-524, 'R', 526-566
A:Cross-references: GB:D10483; GB:J01597; GB:J01683; GB:J01706; GB:K01298; GB:K01990; GB
IDN:BA01334.1; PID:9216483
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A:Reference number: S40531
A:Accession: S40579
A:Molecule type: DNA
A:Residues: 1-22, 'S', 24-126, 'RS', 129-349, 'S', 351-364, 'A', 366-402, 'S', 404-524, 'R', 526-566
A:Cross-references: EMBL:D10483; NID:9216434; PIDN:BA01334.1; PID:9216483
R:Lee, N.; Carbon, J.
Proc. Natl. Acad. Sci. U.S.A. 74, 49-53, 1977
A:Title: Nucleotide sequence of the 5' end of arabid operon messenger RNA in Escherichia
A:Reference number: I41134; MUID:77102763; PMID:189315
A:Accession: I41134
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-14 <RES>
A:Cross-references: GB:K01304; NID:g145312; PIDN:AAA23465.1; PID:g145313
A:Comment: This enzyme catalyzes the phosphorylation of L-ribulose to L-ribulose-5-phosph
C:Genetics:
A:Gene: arab
A:Map position: 1 min
C:Superfamily: ribulokinase
C:Keywords: arabidose metabolism; phosphotransferase

Query Match 14.0%; Score 394.5; DB 1; Length 566;
Best local similarity 28.1%; Pred. No. 1.2e-21;
Matches 166; Conservative 75; Mismatches 221; Indels 129; Gaps 31;

QY 8 IGVNVSQSVRAGIFDLNGLSLSHAT-EKITTTRSGSRVDSQSEIWO--AVSCIRNA 64
DB 5 IGLDFGSDSVRALAVDC-----ATGEEIATIS-----VEMWPR--WQXGQFCOAPRNNQ 49

QY 65 L-----TLADVCAQ--SVAGIGFADTCSL-VVLDKNGDPLPYSPGCD 103
DB 50 FRHHPRDIYESMEALKTVLAEISVEQRAAVVGGVDSGTSTPAIDADGNVLAIRPEFA 109

QY 104 AKON--IIVNMDHRATBOERINATNHPVLN-----YVGGKISPEMERPKILMLKENP 156
DB 110 ENPNAAPFLVMDHRAVEAEETIRLCHAPGVNDYSRTIGGITYSSEWPAKILHYTRDSA 169

QY 157 IYERAGOFFDLADELTWRATG-----DLARSVCVTCKWTWLAHENRWDPDYFTIGL-- 209
DB 170 VAQSAASWIELCDWVPALSLSTTRPDIDIRGRCSAGHSLW--HES-WG-----GLPP 219

QY 210 AELADEDFIRIGHIVSP-----GTPCGNGLTAAQAAEMGLPGTPVAVGLIDAHAG 261

DB 220 ASFDELDPIINRHLPSLFDFTWTADIPVGT-LCPFWAQRILGPESVVISGAFDCHMG 278
QY 262 GIGTGVGEGALNNLAAYEGTSSCTMASTSPSF-----VPGWGPYYSAMVPGMLVVG 317
DB 279 AVGA-----GAQPNMLVKYIGTSTCDILIADKOSGERAVKICGQVDSVVPFGIGLENG 334
QY 318 QSA-----AGAAIDLDLHPAVEAREMAQVNOPLPVWLADRLILEKTAQPSD 366
DB 335 OSAFGDIYAMFGVRLGWPLEQLAQHP-----ELKTQINMS-----QKOLLPLAL 378
QY 367 AVALAKG-----LHVPE-FLGNRAPADPHARAVICGLMERDILNLIYIGLCSTG 420
DB 379 TEAMAKNPISLDHLPVYLDWFMFGRRTPNANORLKGITDNLNLTADAPLFFGGLA---ATA 435
QY 421 YGLRQIIDAQTAQGVSKNITVSGG-AGOHPLVROILLADPCGIP---VITQOCCPEVLIG 476
DB 436 FGARATMECFITDQGLAVNNVVALGSIARKNOVIMQACCDVLRNPLQIVASDQC---ALG 492
QY 477 SAITGAVAGNIAPSVGEAMQOF-THVDKYYVP---QERYSLHHRREAY 522
DB 493 AATFAAVAKVHADIPSAQOKMASAVEKTLQPCSBQNRFEQL-YRRYQW 542

Search completed: March 13, 2003, 16:56:01
Job time: 21.5662 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:42:27 ; Search time 10.6093 Seconds

(without alignments)
2087.642 Million cell updates/sec

Title: US-09-802-208B-4

Perfect score: 2811
Sequence: 1 MTIKTVIGVDSGSGSVRAG.....HHRREYAKYQLHTAKLRD 534

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448.5	16.0	560	1 ARAB_BACSU	P94524 bacillus su
2	447	15.9	564	1 ARAB_BACST	Q98468 bacillus st
3	405.5	14.4	565	1 ARAB_ECO57	P58541 escherichia
4	402	14.3	563	1 ARAB_BACDH	Q9LBP3 bacillus ha
5	397	14.1	563	1 ARAB_MYCSM	P08204 mycobacteri
6	394.5	14.0	565	1 ARAB_ECOLI	P58342 escherichia
7	377	13.4	568	1 ARAB_SALTI	P58342 salmonella
8	368.5	13.1	568	1 ARAB_SALTY	P06188 salmonella
9	361	12.8	563	1 ARAB_YERPE	P58543 yersinia pe
10	352.5	12.5	542	1 MP43_YEAST	P53583 saccharomyc
11	340	12.1	545	1 ARAB_STAM	O99457 staphylococ
12	325	11.6	501	1 XYLB_LACPE	P21939 lactobacill
13	323.5	11.5	524	1 GLPK_MOUSE	O64516 mus musculu
14	323	11.5	530	1 YDEV_ECOLI	P77432 escherichia
15	322.5	11.4	499	1 XYLB_BACSU	P39211 bacillus su
16	321.5	11.4	499	1 GLPK_SULSO	P57907 sulfolobus
17	319.5	11.4	502	1 GLPK_PASMU	P57844 pasteurella
18	317.5	11.3	524	1 GLPK_RAT	O63060 rattus norv
19	312.5	11.1	505	1 GLPK_ENTCA	O34153 enterococcu
20	311.5	11.0	523	1 GLPK_HUMAN	Q14409 homo sapien
21	310.5	11.0	554	1 GLPK_HUMAN	P09099 homo sapien
22	306	10.9	484	1 XYLB_ECOLI	O99464 escherichia
23	306	10.9	498	1 GLPK_LACIA	O99464 lactococcus
24	304.5	10.8	494	1 GLPK_PRRAB	O99407 pyrococcus
25	300	10.7	500	1 GLPK_ENTFA	O34154 enterococcu
26	297.5	10.6	502	1 GLPK_HAEIN	P44400 haemophilus
27	296	10.5	502	1 XYLB_LACBR	P35850 lactobacill
28	295.5	10.5	483	1 XYLB_PASMG	P57828 pasteurella
29	295	10.5	496	1 GLPK_BACSU	P18157 bacillus su
30	290.5	10.3	498	1 GLPK_ECOLI	P37877 escherichia
31	290	10.3	495	1 GLPK_THERO	O99409 thermus bro
32	286	10.2	553	1 GLPK_HUMAN	Q14410 homo sapien
33	283.5	10.1	513	1 GNTR_BACSU	P12011 bacillus su

34	282	10.0	503	1 GLPK_PSETO	O87924 pseudomonas
35	279.5	9.9	497	1 GLPK_BACDH	O98468 bacillus ha
36	278.5	9.9	492	1 GLPK_AQUAE	O66746 aquifex aeo
37	276.5	9.8	497	1 GLPK_PRRD	O93623 pyrococcus
38	271.5	9.7	496	1 GLK2_THEMA	O94164 thermotoga
39	271.5	9.7	501	1 GLPK_ECOLI	P08859 escherichia
40	271.5	9.7	502	1 GLPK_CAEEL	Q21944 caenorhabdi
41	270.5	9.6	496	1 GLPK_THEAQ	O99453 thermus aqu
42	270.5	9.6	513	1 GNTR_BACLI	P46834 bacillus li
43	267.5	9.5	508	1 GLPK_MYCLE	O99481 mycobacteri
44	263.5	9.4	497	1 GLPK_RHIME	O86033 rhizobium m
45	261	9.3	501	1 XYLB_LACIA	O99464 lactococcus

ALIGNMENTS

RESULT 1	ARAB_BACSU	STANDARD:	PRT:	560 AA.
ID	ARAB_BACSU			
AC	P94524: 005185:			
DT	01-NOV-1997 (Rel. 35, Last Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	L-ribitol kinase (EC 2.7.1.16).			
GN	ARAB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423:			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168:			
RX	MEDLINE=97124191; PubMed=8969504;			
RA	Wipat A., Carter N., Bignell C.S., Guy J.B., Piper K.,			
RA	Sanders J., Emerson P.T., Harwood C.R.;			
RT	"The dnaB-pheA (256 degrees-240 degrees) region of the Bacillus			
RT	subtilis chromosome containing genes responsible for stress			
RT	responses, the utilization of plant cell walls and primary			
RT	metabolism."			
RL	Microbiology 142:3067-3078(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168:			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Berttero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Bouriss R., Bourisier L., Brans A., Bron S.,			
RA	Brouillet S., Brunsch C.V., Caldwell B., Capiano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullio M.F., Iraya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogihara A., Ondega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porrolik S., Prescott A.M.,			
RA	Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			

RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viart A., Wambuit R., Wedler E., Wedler H., Weitzengruber T.,
 RA Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RN [4]
 RP TRANSCRIPTIONAL REGULATION.
 RX PubMed:10417639.
 RA Mota L.J., Tavares P., Sa-Nogueira I.M.G.;
 RT "Mode of action of Arar, the key regulator of L-arabinose metabolism
 RT in Bacillus subtilis.";
 RL Mol. Microbiol. 33:476-489(1999).
 CC -1- CATABOLIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
 CC phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- INDUCTION: Transcription is repressed by glucose and by the
 CC binding of arar to the operon promoter. L-arabinose acts as an
 CC inducer by inhibiting the binding of arar to the DNA, thus
 CC allowing expression of the gene.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X89408; CAAG586.1; -
 DR EMBL: Z75208; CAAG9588.1; -
 DR EMBL: Z99118; CAB14839.1; -
 DR Subtilase, Bgl1905, arab.
 DR Pfam: PF00370; FGGY; 1.
 DR Pfam: PF02782; FGGY_C; 1.
 DR TIGRFAMs: TIGR01234; L-ribulokinase; 1.
 DR Transferase: Kinase; Arabinose catabolism; Complete proteome.
 KM CONFLICT 59 P -> F (IN REF. 1).
 FT SEQUENCE 560 AA; 61009 MW; DE721AB73AF0AEF CRC64;
 SQ
 Query Match 16.0%; Score 448.5; DB 1; Length 560;
 Best Local Similarity 27.1%; Pred. No. 4,4e-25;
 Matches 153; Conservative 95; Mismatches 237; Indels 79; Gaps 20;

QY 341 ----MAQVNOPLPWLAADRIIEKTAQSPDAVALAKGLHVPEFLGNRAPFADPHARAV 395
 DB 358 GVHLLSEKKNHQAQPG-----ESGLALDMWNNRSTLVADADLTGM 398
 QY 396 ICGIGMERDLDNLALYIAGICGIGYLRQILDQOTAGVSKNIVISG-AGCHPLVRQ 454
 DB 399 LIGWTL--LTKPEIRALVEATVYCTRMIEFKESGVIEELFAAGIAEKNPVMQ 455
 QY 455 ILADTCIPVITTCOCPEVLGSAIIGAVG-----NTAPSVGAMQGFTHVDK 503
 DB 456 IYADVTMDIKISGSPAPALGSNIFGALAGKKGKGYDDIKKAANMGLK-----DI 509
 QY 504 VYPOERYOSLHRRYEAQKQLOH 527
 DB 510 TYTPNAENAAVEXKLYAEKELVH 533
 RESULT 2
 ARAB_BACST
 ID ARAB_BACST STANDARD; PRT; 564 AA.
 AC Q95468;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 OS ARAB.
 GN Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-T-6;
 RA Gillee-Gropper S., Shoham Y.;
 RT "The L-arabinose utilization gene cluster from Bacillus
 RT stearothermophilus T-6.";
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATABOLIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
 CC phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF160811; MAD4517.1; -
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY; 1.
 DR Pfam: PF02782; FGGY_C; 1.
 DR TIGRFAMs: TIGR01234; L-ribulokinase; 1.
 DR Transferase: Kinase; Arabinose catabolism.
 KM SEQUENCE 564 AA; 62202 MW; DBFPDBFC36EDB5 CRC64;
 SQ
 Query Match 15.9%; Score 447; DB 1; Length 564;
 Best Local Similarity 28.5%; Pred. No. 5,7e-25;
 Matches 159; Conservative 89; Mismatches 245; Indels 64; Gaps 20;

Qy	165	FDLADFLTWRTGGLASVCTVVCCKMTWLHENRMDPDYERTTG--LAELADDFIRIGH	222
Db	182	LEGTDWYIFKKTGTVIVNCSCAAGTSGTWHKODGPKSEFFRALDPREHTLTE---TKLKG	238
Qy	223	HIVSPGTPCGNGLRPAQAAEKGLLPGTPVAVGLIDAHAGGIGTVGEGGALNNMLAYVFGT	282
Db	239	PIVPLGTRAG-VLTLENMAAMMGLLPGTAVAVGVNDAAHAPGVGVPEPG---KLVMAMGT	294
Qy	283	SSCTMASTSPSEVPVGVWGPYYSAMVPGMLVBEGOSAGACAIQDLDPHAYEE-----	337
Db	295	SICMHLLETBEKYVEGCKGVEDGIIIPGYGYEAGOSAVGDI-----FAMVYEGGVAPY	348
Qy	338	AREMAVQNQPLPVWLADRLLEKTA---QPSDAVALAKGHHVPEFLGNAPRADPHARA	394
Db	349	VKEAEKEGSGISVHEM-----LEKRAAYRPGET-----GLLALDMMNGNSVLYPDTDLTG	398
Qy	395	VICGLGERDLDNLALYIAGLCIGYGLRQIIDLDAOTGAQGVSNKNYISGAGQ-PLVLR	453
Db	399	LIVGYTL---LTKPEELIYRALLEGTATGTRKIIDAFENVGVKKDELACGGLRQXKKLLM	455
Qy	454	QILADTCGIPYITTOCCPEVLLGSAILIGAVAGNIA---PSVGEAMQOFTHV-DKYYYPQ	508
Db	456	QIYADVTVNRKIIKIASKOTPAVGCAMPAFAVAAGKENGYESIVEAARNMGKVARETFKPI	515
Qy	509	ERYOSLHRRRYEAQKOL	525
Db	516	PENVAMTEOLYQOETTKL	532

RESULT 3
 ARAB_ECO57 ID ARAB_ECO57 STANDARD: PRT: 565 AA.
 PS8541:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR 20072 OR ECS00657.
 OS Escherichia coli O157:H7.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posifal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocistus K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1 CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
 CC phosphate.
 CC -1 PATHWAY: L-arabinose catabolism; second step.
 CC -1 SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its use

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/
CC or send an email to%5Blicense%5B-sib.ch%5D](http://www.isb-sib.ch/announce/or_send_an_email_to%5Blicense%5B-sib.ch%5D)).

CC
DR EMBL; AE005183; AAC54367.1; -
DR EMBL; AP002350; BAB33450.1; -
DR TIGRFAMs; TIGR01234; L-ribulokinase; 1.
KW Transferrase; Kinase; Arabinose catabolism; Complete proteome.
FT INT_MET 0 By SIMILARITY
SQ SEQUENCE 565 AA; 61096 MW; 45E7D9AA096B9CAD CRC64;

	Query Match	14.4%	Score 405.5	DB 1	Length 565
	Best Local Similarity	27.7%	Pred. No. 5,8e-22		
	Matches 163	Conservative 77	Mismatches 224	Indels 125	Gaps 29
QY	8	IGVDVSGSVRAGIFDLNGLSLSHATEKITTTTRNSRFEVSOSSFIOW--AVSCSRNAL	65		
Db	4	IGLDFGSDSVRALVYD-----CTTAGEIATSVEMYPR--WQKGQFCAPANNQF	49		
QY	66	-----TLADYCAQ---SVAGIGFDATCSL--VYLDKKGDDLPVSPGDA	104		
Db	50	RHHPRDYIESMEALKYTLAEISVEQRAAVVIGVDSTGSPAPADAGNALALRPEAE	109		
QY	105	KQN--IIVMDHRAEQEARNATHHPVLN-----YVVGKISPMEHTPKILMLKNMEPI	157		
Db	110	NPNAFVFLMKDHTVAEEAEETRLCHAGVNDVSRYSIGGISSEWFMALKILHYTRQDSAV	169		
QY	158	YERAGQFDDLADFLTWKATG-----DLARSCYTYCKMTWLAHENEKMDPYFRITGL--A	210		

Db	170	AQSDASWLELDWV	PALLSGITTRKQDKSGHSLW	-HES-WG-----	GLPPA	216
Qy	211	ELADEDEFRIGHHNVSP	-----GTPCCNGSLTAQAAMEMGLPGHPVAVGLIDAHAAG			262
Db	220	SFFDELDPILNRHL	PSPLFTDWTADIPVGT	-LCPEMAQRLLGPRESVISGAFDCMHGA		278
Qy	263	IGTVGVEGAGLNANL	AYVFGTSSCTMASTTSPSF	-----VPGVMPTYSAMVPGMLVLEGGQ		318
Db	279	VGA----	GAQPNALVVKVYISTCDIL	ADKOSGERAVKIKISQVDGSSVPGTGLEAQ		334
Qy	319	SA-----	AGAAIDOLLDFHPAVEAREMAQRVNOPLP	-----VWLADRILEKTAQ		363
Db	335	SAFGDIYAMFGRVLTGMPLEOL	AAQHPRLKIQIDASK	--OLLPALTEAMAKNPSLDH---		389
Qy	364	PSDAVALAKGHVPE	-FLGRRAPFADPHARVYCGSGMERDIDNLALYINAGLSIGYG			422
Db	390	-----	LPVVLDMPNGRRTPNAQRILKGVTTDNLITDAPLFGCLLA	-----ATAFG		436
Qy	423	LRLQILDQTAQGVVSKNIVISG	-AGOHPLVRQLLADYTCGP	-----VITTPQCCPEVLLGA		478
Db	437	ARAINECTTDGILVNNVMAL	GGIARRKNQYIMQACCVLRLRPQIYVASDCC	-----ALGAA		493
Qy	479	ILGVAVAGLIAPSGVEAMQOF	-THVDKLYVVO----	ERYOSLHRRYEA		522
Db	494	IFAAVAKVHADIPSAQDKMAS	AVEKTLDPRESEQQNFQDL	-YRRTIQM		541
RESULT 4						
ARBAB_BACHD	ARBAB_BACHD	STANDARD:	PRT:	563	AA.	
ID	Q9K8O3:					
DT	15-JUN-2002 (Rel. 41, Created)					
DT	15-JUN-2002 (Rel. 41, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	L-Ribulokinase (EC 2.7.1.16).					
GN	ARAB OR BH1872.					
OS	Bacillus halodurans.					
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.					
OX	NCHI_TaxID=86665;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=C-125 / JCM 9153;					
RX	MEDLINE=20512582; Pubmed=11058132;					

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001513; BAB05591.1; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY_1.
DR TIGRfams: TIGR01234; L-ribulokinase; 1.
DR TIGRfams: TIGR01234; L-ribulokinase; 1.
KM Transferase: Kinase: Arabinose catabolism: Complete proteome.
SQ SEQUENCE 563 AA; 61668 MW; CCBC39006669ACD CRC64;

Query Match 14.3%; Score 402; DB 1; Length 563;
Best Local Similarity 26.7%; Pred. No. 1e-21;
Matches 155; Conservative 82; Mismatches 260; Indels 84; Gaps 20;

QY 4 TKYIVGVDSGSVRAGIFDL-NG-SLSHATE-----KITTRRGSGRVESSQ 51
DB 3 TKYIVGVDSGSVRAGIFDL-NG-SLSHATE-----KITTRRGSGRVESSQ 51
QY 52 EIMQAVSCIRNALTLADVCASVAGIGFDP-CSLVLDKNGDPLRPSPEGDAKON-- 107
DB 63 DYVEVLITSPRAVVKESGVADDDYIGIGVDPFTACTMLPVDESGPCLLQA--YKDNPHS 120
QY 108 -IIVMDHRAATEQAEIRIN---ATHNPVLNVGKISPEMETPKIILKEMPEITERAG 162
DB 121 VVKLMKHNHAADKANAINEMAERKGEAFLLPRYGKISSEWIAKVMQIILDEADVDVNRTD 180
QY 163 QFEDLADFLTRARAGDLARSCVTCTCKWTMLAHENRMDPROFRTIGLAELDADEFIRGH 222
DB 181 QFLEATMIVISQMTGKIVKNSCTAGYKAIMHKRGYPSNEFFKALD-PRLEHLLTTTKLRG 239
QY 223 HIYSPGTPCGNGLTAAQAAEMGLPSTPAVAGLIDAHAG---GIGTVGEGALNNLAY 278
DB 240 DIYVLGERAG-GILPEMAEKGLNPGIAVAVGANDAHAAVPAVGVTPRG-----KLYM 291
QY 279 VFGTSSCTMASTTSPSFVPGVWGPYSAWPGMLVEGGSAGAAIDQLDFHPAY--- 335
DB 292 AMGTSICHHMLGEKEQEGVCEGVEDGIIIPGYLGVEAGSAGVADIPAMFVKHVSATF 351
QY 336 EEARERMAQRNQLPVLADRIIEKTAQ--PSDAVALAKGIHVPELGNRAPADPHAR 393
DB 352 DEAEKGVNVHALL-----EKASQLRGEES-----GLLADWMNGNSIIVDELIS 398
QY 394 AVICGLGEMERDLNLALYTAGLGIGYGLROIIDAQTAQGVSKNIVISGAGQ-HPV 452
DB 399 GMLGTYLQTKRPE---TYRALLEATFAGTRAIADAFHRCGVEVHELYACGGLPQKNHLL 455
QY 453 ROIILADTCGIPVITTOCEPVLGSAIIGAVAGNIA---PSVEAMQOF----- 498
DB 456 MQIADVTNEIKVAASKQRPALGAAMFASVAAAGSEVGVDISEAKKMGKRVADENFRK 515
QY 499 --THVDKTYTPQERTYSLH-----RRYAYVQLOH 527
DB 516 IPEHVAITEKLYOEYVTLHDYFGRGANDVVKRLKALKSIQH 556

RESULT 5

ARAB_MYCSM
ID ARAB_MYCSM STANDARD; PRT; 563 AA.
AC Q9LBO3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE L-ribulokinase (EC 2.7.1.16).
GN ARAB.
OS Mycobacterium smegmatis.
OC Bacteria: Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMDU;
RA Soda A., Takata G., Izumori K.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB038527; BAB92524.1; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY_1.
DR TIGRfams: TIGR01234; L-ribulokinase; 1.
KM Transferase: Kinase: Arabinose catabolism.
SQ SEQUENCE 563 AA; 60432 MW; 12F097597DD07918C CRC64;

Query Match 14.1%; Score 397; DB 1; Length 563;
Best Local Similarity 29.3%; Pred. No. 2.4e-21;
Matches 166; Conservative 64; Mismatches 249; Indels 88; Gaps 19;

QY 5 KTVIGVDVSGSVRAGIFDL-NGSLSHATEKITTRRSRVE----- 47
DB 8 KYVIGVDPFTLSGRALVRAVDGAEKSAER---TYRHGVTEALDPREVRLLPADYAL 63
QY 48 QSSQEIQAVSCIRNALTLADVCASVAGIGFDP-CSLVLDKNGDPL-----PVSP 100
DB 64 QVRADYIDLRFALPEAVANAGIDPADVVGIGTDFCTACTVAAATSDGTPICQDEFAADRP 123
QY 101 EGDAKONIIVMDHRAATEQAEIRIN---ATHNPVLNVGKISPEMETPKIILKEMPE 156
DB 124 HAYAK---LWRHNHAOPQADRINALAAARGETMLPRYGGLISSEWEFAKGLDLEDEPE 179
QY 157 IYERAGQFEDLADFLTRARAGDLARSCVTCTCKWTMLAHENRMDPROFRTIGLAELDA 214
DB 180 IYAADRWVSGADWIVQLTGRYVRNISTAGYK-----AIRDQKYPSPAFELMPGF 233
QY 215 ----EDFIR--IGHNIYSPGTPCGNGLTAAQAAEMGLPSTPAVAGLIDAHAGIGTVGV 268
DB 234 ASFYEDKITEQPIGLGFAAGT-----LTAAAMATGIPREGIIVAVGNDIHNHVAADADA 288
QY 269 EGCALNNLAYVFGTSSCTMASTTSPSFVPGVWGPYSAWPGMLVEGGSAGAAIDQL 328
DB 289 DPG---RLAIIMGTSTCHVNMGRFLRVPCMGVGDGITGLMGVBAQSGVDIPAMF 345
QY 329 LDFHPAVEEARERMAQRNQLPVLADRIIEKTAQOPSDAVALAKGIHVPELGNRAPFA 388
DB 346 TKNCVPKEIAIEASRR-----GLTAEHLSLAEDQEVG---EHGLVALDMHSGNRSVLV 397
QY 389 DPHARAVICGEMERDLNLALYTAGLGIGYGLROIIDAQTAQGVSKNIVISGAGQ 448
DB 398 DHHLSGIMVGOTLD---TTCVDDYTRALLEATFAGTRAIIVETFORSGVPELVLVAGGLIK 454

ID ARAB_SALTY STANDARD: PRT: 568 AA.
 AC P58342;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR STY0120.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_Taxid=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18.
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parikh H.J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahina M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagers K.,
 RA Krogh A., Larsen T.S., Leather S., Moulie S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrrell B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose -> ADP + L-ribulose 5-
 CC phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL627265; CAD01260.1;
 DR TIGRFBAS; TIGR01234; L-ribulokinase; 1.
 KM Transferase; Kinase; Arabinose catabolism; Complete proteome.
 FT INIT_MET 0
 SO SEQUENCE 568 AA; 61643 MW; A1515DB7249A0F75 CRC64;
 Query Match 13.4%; Score 377; DB 1; Length 568;
 Best Local Similarity 26.4%; Pred. No. 6,8e-20;
 Matches 151; Conservative 82; Mismatches 249; Indels 90; Gaps 24;

QY 324 --AIDQLDHPRAVEAREKARVNOPLPWADRIIEKTAQPSDAVALAKGLHWPE-- 379
 Db 349 SWPLEQIAQHPBELKTIINSQK--QLLPAL-----TDAMAKRPSLDHLPVL 394
 QY 380 --FLGNAPAPADPARAVITGICMERDLDNLALYIGICIGVGLQIIDAOAGVGS 437
 Db 395 DWNGRTPNPNORLKGVIYIDLNLATAP--ALFGGLVSTAFGARAIQECFTGQIAV 451
 QY 438 KNIVISGC-AGQHPVLVQIADTCGIP--VITQCCPEVLLGSAILGAVAGNIAPSGE 493
 Db 452 NNWVALGIAKKNQVIMQVCCDVLNRPDLQIVASQCC--ALGAAITPAVAAKVHADIPA 508
 QY 494 AMQGF-THVDKYYTPQERTQSLHHRREAYKQ 524
 Db 509 AQDSMASAVERTLRPREDAQRFERYRYQQ 540
 RESULT 8
 ID ARAB_SALTY STANDARD: PRT: 568 AA.
 AC P06188;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR STY0103.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85232044; PubMed=2989100;
 RA Lin H.-C., Lei S.-P., Wilcox G.;
 RA "The arabid operon of Salmonella typhimurium LT2. I. Nucleotide
 RT sequence of arab and primary structure of its product,
 RT ribulokinase";
 RL Gene 34:111-122(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Portolick S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose -> ADP + L-ribulose 5-
 CC phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M11047; AAA27023.1;
 DR EMBL: AE008698; AAL19067.1;
 DR PIR: A24984; A24984.
 DR StGene: SG10013; arab.
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY_1.
 DR Pfam: PF02782; FGGY_C_1.
 DR TIGRFBAS: TIGR01234; L-ribulokinase; 1.
 KM Transferase; Kinase; Arabinose catabolism; Complete proteome.
 FT INIT_MET 0
 SO SEQUENCE 568 AA; 61643 MW; A1515DB7249A0F75 CRC64;
 Query Match 13.4%; Score 377; DB 1; Length 568;
 Best Local Similarity 26.4%; Pred. No. 6,8e-20;
 Matches 151; Conservative 82; Mismatches 249; Indels 90; Gaps 24;

SQ SEQUENCE 568 AA; 61620 MW; B1D838C37B7C4134 CRC64;
 Query Match 13.1%; Score 368.5; DB 1; Length 568;
 Best Local Similarity 26.8%; Pred. No. 2.8e-19;
 Matches 156; Conservative 79; Mismatches 235; Indels 113; Gaps 27;
 8 IGVNDSGSGVRAGIDFDLNGSLSLSHAT-EKITTTRSGSHVEOSS-----50
 4 IGLDGSQSDVRALAYDC-----ATGDEIATSVEMYPWMQECRYCDGPNNOFRHHPRDY 56
 51 QEIMQAVSCINNALTLADYCAOSVAGIGFDATCSL-VLIDKNDPLVPVSPGDAKON--107
 57 MESMEALAKAVLAQLSAQA--RANVYIGVDSTGSTPAIDAGNVALRPEFAENPM 114
 108 IIVMMDHRTAEQERINAHN-----PVLNVYGGKTSPEMERPKTLMLKENPEIYERAG 162
 115 EYLMKDHRTAEADEITRLCHPRKVDYSRYIGIYSSSEFMAKLILHVRQDSAVQAQAV 174
 163 QFEDLADFLTWRTATG-----DLARSVCVTCKMTWLAHENRMDPDYFRTIGL--AEIAD 215
 175 SWIELCDWVPALLSGTTRPQDIRGRCSAGHKTLM--HES-WG-----GLPRAFPDE 224
 216 DFRIGHNIVSP-----GTPCCNGLTQAQAAEMGLLPFPVAVGLIDAHAGIGTVG 267
 225 LDPICINRLRYPLFSETFTADLPVGT-LCAEWAQRILDESVYISGAFDCHMGAVGA--281
 268 VEGGALNNLAVYFGTSSCTMASTTSPSF-----VPGWVGYSYAMVNGMLVBEQGSAAGA 323
 282 --GAQPNILVKYIGTSTCILADKOSVGDRAVKLGIGVDGSDVNFETGLEQGSAGFD 339
 324 -----AIDQLDFPAAVEAREMAQRYNOLPFWLADRLILEKTAQPSDAVALAK 372
 340 IYAMFSRVLSWPLEQLAQHPELKQINASOK--QLLPAL-----TDAAKNP 385
 373 GLHYVE----FLGNRAPADPHARAVICGLMERDLMLALYIAGLIGYGLRQIID 428
 386 SLIDHLPVLDWENGRTRPANORLKGVIYDMLATDAP--ALFGLVASTAGARADE 442
 429 AQTAGGVSNKNIVISGG-AGOHPLVROLIADTCGIP--VITTOCCPEVLGSAIIGAVA 484
 443 CTDSQIAGANNMALGIAKKNOVIMQCCDVLNRPLOIVASDQCC--ALGAATIAAVA 499
 485 GNIVDSVGEAMQOF-THVDKYYYPQ---ERYQSLHRRYEAV 522
 500 AKVHADIPAAQOSMASAVERTLPRHPEQAQRFQQL-YRRYQOM 541
 RESULT 9
 ARAB_YERPE STANDARD; PRT; 563 AA.
 ID ARAB_YERPE P58543;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR YP02254.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=632;
 OX NCBI_TaxID=632;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Blovat Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebalnia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holtroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";
 Nature 413:523-527(2001).

CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-phosphate.
 CC -1- PATHWAY: L-arabinose catabolism; second step.
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: AJ414151; CAC91058.1; ALP INT.
 DR TIGR: TIGR01234; L-ribulokinase; 1.
 KW Transferase; Kinase; Arabinose catabolism; Complete proteome.
 SQ SEQUENCE 563 AA; 61833 MW; 88FE1EB00A5414E CRC64;
 Query Match 12.8%; Score 361; DB 1; Length 563;
 Best Local Similarity 26.4%; Pred. No. 9.7e-19;
 Matches 152; Conservative 81; Mismatches 257; Indels 86; Gaps 25;
 8 IGVNDSGSGVRAGIDFDLNGSLSLSHATEKITTT-----RRSGRVEQSSOEIMQ 55
 10 IGLDFGSDSVRLAYD-----CONGTEIDTEVYVPRMKKGLYCAQAOQFRHHPLDYIE 64
 56 AYCSCIRNML-PLADYCAOSVAGIGFDATCSL-VLIDKNDPLVPVSPGDAKON--IIVW 111
 65 ASEQAIRQWKRILSEBORHIVGIGVDSGSTPAPRIDEQOVLALPDRADNPAMFVLM 124
 112 MDHRTAEQERIN-----ATHHPVLNVYGGKTSPEMERPKTLMLKENPEIYERAGQFDP 166
 125 KQHTALEEAEINRLORSEGFADYSRYIGYSSSEFMAKLILHVTADVAVRAAASWIE 184
 167 LADFLTWRTATG-----DLARSVCVTCKMTWLAHENRMDPDYFRTIGLAEIAD 221
 185 LCDWVPALLSGTTPADODIRGRCSA-----GHKSLMPSWGLPRAFLALDYSLVN 237
 222 H-----HIVSGTCCNGLTQAQAAEMGLLPFPVAVGLIDAHAGIGTVGEGGALN 274
 238 DLDYPMFTDTYTAERYVQO-ITAEWAERLGLPTTVILSGAFCHMGAVGA---GAQPY 292
 275 NLAVYFGTSSCTMA-----STTSPSFVPGWVGYSYAMVNGMLVBEQGSAAG---AAID 326
 293 TLVKYIGTSTCIDLADDDQVGRGRLAIGICQYEGSVLG-WTGMEGQSAFEDWYAMFS 351
 327 QLIDF--HPAVEARMAQRYNOLPFWLADRLILEKTAQPS-DAVALAGLHYVPEFLN 383
 352 NLISWPLHQAALTOPRPMOQLKQIESNLASLTRAQAONPSIDHLPV---VLDFNGR 406
 384 RAFPADPHARAVICGLMERDLNMLALYIAGLIGYGLRQIIDQATAGGVSNKNIVIS 443
 407 RFPNANORLKGVIYDMLATDAPFLFGFLA---ATAFGRLATBECFEOODIPIDNVLAL 463
 444 GG-AGOHPLVROLIADTCGIP--VITTOCCPEVLGSAIIGAVANINAPVGEAMQOFT 499
 464 GGIARSPVIMQCADVMNRPLDIVASDQCC--ALGAATIAVAAGAHDDVPTAQRHMA 520
 500 -HVDKYYYPQ---RYQSLHRRYEAYKOLHTAK 530
 521 CNIEFTLIPDPVQVYRYQQL---YORYQOMCHTAE 552
 RESULT 10
 MP43_YEAST STANDARD; PRT; 542 AA.
 ID MP43_YEAST P53583;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MP43 protein.
 GN MP43 OR YN1249C OR N0875.
 OS Saccharomyces cerevisiae (Baker's yeast).

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN SEQUENCE FROM N.A.
RC SRRAIN-5286C / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeher U., Belinhaber J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
RN
RN SEQUENCE OF 41-542 FROM N.A.
RC STRAIN-ENY.WA-1A;
RA Boles E., Hettmann C., Zimmermann F.K.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X96722; CAAG5495.1; -
DR EMBL: 271525; CAAG6156.1; -
DR EMBL: X94214; CAAG6905.1; -
DR SGD: S0005193; MPA43.
DR CONFLICT 46
FT SEQUENCE 542 AA; 61667 MW; 830A8B4E95365AD CRC64;
SQ
Query Match 12.5%; Score 352.5; DB 1; Length 542;
Best Local Similarity 25.9%; Pred. No. 3.8e-18;
Matches 127; Conservative 78; Mismatches 190; Indels 95; Gaps 22;
QY 8 IGVGVGSGVRAGIFD-LNGSLSLSHATEKTTTTRSGS---RPEQSOEIMQAVSCIRN 63
DB 9 IGVGVGSSARIGVNYNYNDALLEMAQEPVYODSSKSKWKEWKEIIRKALOKCLOK 68
QY 64 ALTLADYCAQSVAGIGFDATCSLVVL--DKNGDPLVSPGDAQNTIIMMDHRTAQAE 121
DB 69 -----LNIREVSKSGVSAFCSLAIFERDRTSNMLIRPNED---NIFWDSAVAVECQ 121
QY 122 RIN-ATHHPVNLVYGKISPEMETPKILMLKENNPEIYERAGOFFDLADELTWRATDGLA 180
DB 122 WLNMQCPQQLLDYLGKRFVPEMGVPLKFLDEYSHLRDKNFHIFDLHYIAY----ELS 177
QY 181 RSVCTVYCKTW-----LAHEN-----RMDPDYFRTIGLAELADEDFRIGHNI 224
DB 178 R-----LEWNIIEGLLGRNGLNGICNDGEVSGWSSSFYKNI-----INLPSN- 219
QY 225 VSPGTPCGNGLTQAQAAEMGLLPCTPVAAGLIDAHAG--GIGTVGVEGALNNLNAVYEGT 282
DB 220 VSIQGT---TSLVANKH-----STIYVRSCIDISYASWFAVASHPLE-----TSLPMTAGT 266
QY 283 SSGTMASTT-SPEFPGVWGPY-----YSAMVPGMLVLEGQSAAGAIDOLL 329
DB 267 SSCGMVGTITSDIRIPGVWGPEDTIIIDNRGDFSYAA-----GQSGTGLIHLEF 316
QY 330 DHPHVAEAREMAQRVQPLPWLADRILEKTAQPSDAVALAKGLHVPFLGRAPAD 389
DB 317 ESHHCARKIILKDGADIVQVLEQFTRD--IREKNNGLSTIHI-LTKDMFEYGYEGNRTFPAD 373
QY 390 PHARAVICIGMERDLDNLALYTAGLGIGYGRQILD--QTAQGVVSKNIYISGAG 447
DB 374 PRIGSGTGSTSTSMNLTYKYICILEFSLFQTKLIIDTFQENSNHIIKEKISGQA 433
QY 448 QHPLVROILA 457
DB 434 KNERLISLIS 443

```

```

RESULT 11
ID ARAB_STAM STANDARD; PRT; 545 AA.
AC 099W57;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE L-ribulokinase (EC 2.7.1.16).
GN ARAB OR SAV0552 OR SA0510.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN
RN SEQUENCE FROM N.A.
RC SRRAIN-Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kenehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
CC phosphate.
CC
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC
CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AP003359; BAB56714.1; -
DR EMBL: AP003130; BAB41741.1; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY_1.
DR Transferrase; kinase; Arabidose catabolism; Complete proteome.
SQ SEQUENCE 545 AA; 60965 MW; E3917036237CEA4B CRC64;
Query Match 12.1%; Score 340; DB 1; Length 545;
Best Local Similarity 24.7%; Pred. No. 3.1e-17;
Matches 140; Conservative 109; Mismatches 245; Indels 72; Gaps 23;
QY 8 IGVGVGSGVRAGIFD-LNGSLSLSH-----ATEKTTTTRSGSRVE-----OSSOEIMQAV 57
DB 5 IGVGVGASGRVFLINTNGQVSKFVPRYTHGVISELNGKLRPHYALQNSNDYLEIM 64
QY 58 CSCIRNALTLADYCAQSVAGIGFDATCSLVVL--DKNGDPLVSPGDAQO-----NI 109
DB 65 EGGISYIVRSKIDPVNIVGIDFTSTIIFDEN-----LNVHNLKQFKNPNHAYVK 119
QY 110 VMMDHRTAEQERINAT---HHPVLVYVGKISPEMETPKILMLKENNPEIYERAGOFF 165
DB 120 LMKHHGAYKEAEKLYQTAIENNNKWLCHYGYNVSSSEMMIKRIMEVNRARPEIMEKTYIM 179
QY 166 DLADFLWRATGDLARSVCTVYCKTWLAHENRMDPDYFRTIG--LAELADEDFRIGHN 223
DB 180 EAGWIVNKLTKNNKVRNCGEFAFW--EEETGFHYDLFPKIDPKLSKVIOD---KVSAP 235
QY 224 IVSPGTPCGNGLTQAQAAEMGLLPCTPVAAGLIDAHAGIGTVGVEGALNNLNAVYEGT 283
DB 236 VVNIQGEVYVK-LDDKMKQKIGLSKETMVSPIIIDAHASLLG-IGSEKD--KEMTMVWGT 291

```

Db 1 MSAAVLGIDIGTSAVAVKSAIDKDGNNVVAQASAKYALADQPHPGYSQDPEDEWVOTTAIR 60
 QY 63 NALLTADYCAQSAVAGIGIGDATC -SLVYLDKNGDPLPVSPEDGAKONIIVMDHRTQDAE 121
 Db 61 ELLQOGEVTAQDIEGLSISYSGOMHGLVLLDESAATV -RPAIIMNDTRTSSOCR 111
 QY 122 RINATH -HPVANTYVGKISPEMETPKILMEKENPEIERAGQDFDLDFLTRWRTGSLA 180
 Db 112 ELESQFGDDFKITIGNRPLEGFTLPKLLWVENEENPINKRAFTFLPDLYLRRTGKLA 171
 QY 181 -RSVCT-----VTCMTWLAHENRMDPDYFRTIGLAELADEDFIRIGHIIVSPGTPC 231
 Db 172 MDRSDATGTVLDDITTSQWSEFLCNQDIP-----LFLCPPLISTAVVGH----- 217
 QY 232 GNGLTQAQAAMGGLPGTPPAVGLIDAHAGIGTGVGCGALNNIAY -FGTSSCTMA-S 289
 Db 218 ---INQTAQLSGLSVNTKVFEGADNAGAVGA---GILSDKALVSGTGVLLKYE 270
 QY 290 TTSPSPFVGWVGPPYSAAVPGMLVEGQSAAGAIQDL-----DFHPAVEAREMA 342
 Db 271 DNAQDVRGLV-QYERHAPCKRTYSMGVTLAAGTSLNFKOTFPAPDEDFGVVASAEQ-- 327
 QY 343 QRVAQPLPFWMLADRILEKTAOPSDAVALAKGLHVPEFLGNRAPADPHARAVICGL-GM 401
 Db 328 -----STIGANGTLFPAYIVGEAPAPADATIRGSPFGVGS 363
 QY 402 ERDDNLTALYTAGLCIGYGLRQILDQAOGVVSKNIVISGAGQHPVLRQLADTCG 461
 Db 364 HQRADFVRAV---LEGITFFSEFDILIKYOHNGAEFKTIVSIGGAKSALMLQIADIFN 419
 QY 462 IPVITTOCCEVVLGGSALTGAVAGNIAPSVGAMQGFHVQKYYXPOERYOSLHRRREA 521
 Db 420 CKVVSLEKNEGPGMGAMAIATATGAGWEKFTLADCAQTFVHYGKATPYTAHAQYQEMRL 479
 QY 522 YKOL 525
 Db 480 YQOI 483
 RESULT 13
 GLPK_MOUSE
 ID GLPK_MOUSE STANDARD; PRT: 524 AA.
 AC 064516:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GlyceroL kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase)
 DE (glycerol kinase) (GK).
 GN GK OR GYK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090; 1
 RN 1)
 RP SEQUENCE FROM N.A. 1
 RC STRAIN-BALB/c;
 RX MEDLINE=97038697; PubMed=8884278;
 RA Hug A.H., Lovell R.S., Sampson M.J., Decker W.K., Dinulos M.B.,
 RA disteche C.M., Craigen W.J.;
 RT "Isolation, mapping, and functional expression of the mouse X
 RT chromosome glycerol kinase gene.";
 RL Genomics 36:530-534(1996).
 CC -1 FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
 CC METABOLISM (BY SIMILARITY).
 CC -1 CATALYTIC ACTIVITY: ATP + glycerol = ADP + glycerol 3-phosphate.
 CC -1 PATHWAY: glycerol utilization, rate-limiting step
 CC -1 CYTOPLASMIC LOCATION: BOUND TO THE MITOCHONDRIAL SURFACE OR
 CC -1 SUBCELLULAR (BY SIMILARITY).
 CC -1 SIMILARITY: BELONGS TO THE FPKOKINASE / GLUCOKININASE /
 CC GLYCEROKINASE / XYLUKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U48403; AAC52824.1; -
DR HSSP; P08859; 1GLJ.
DR MGD; MGI:106594; GYK.
DR InterPro; IPRO00577; FGGY_kln.
DR Pfam; PF00370; FGGY; 1.
DR Pfam; PF02782; FGGY_C; 1.
DR TIGRfam; TIGR01311; glycerol_kln; 1.
DR PROSITE; PS00445; FGGY_KINASES_2; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 1.
KM Glycerol metabolism; Transferase; Kinase; ATP-binding.
FT NP_BIND 167 179 ATP (PROBABLE).
SQ SEQUENCE 524 AA; 57458 MW; C3E1C56887683B55 CRC64;

Query Match 11.5%; Score 323.5; DB 1; Length 524;
Best Local Similarity 25.6%; Pred. No. 4.6e-16;
Matches 133; Conservative 74; Mismatches 220; Indels 93; Gaps 20;

OY 5 KTVIG-----VDVGGSVRAGIFD-LNGSLSHATEKITTRRSGSVESQSIQAVC 58
DB 6 KAVIGPLVAVDOCTSTRTFLVNSKTAELSHHOVEIKOEFPEGWEODPKELDSYV 65
OY 59 SCIRNA---LTLADVCAQSVAGIGF-DATCSLVVLDK-NGDPLFVSPEGDAKQNIYMD 113
DB 66 ECIEKTECEIGQLNIDISINIKAIQVSNORETYVMDKVTGPEL-----YAAVWLD 116
OY 114 HRATEQAE---RINATHHPVLNVGCKISPEMETPKILMKEMPELYERAGO----FF 165
DB 117 LRQSTVENLSKRIIPGNNNFYKSTGLPLSTYFSAVLRMLDVKVQVEAEEENRALFG 176
OY 166 DLADFLTRATGDLARSY-CT--VTCKWTML--AHENRMDPYFRTIGLAELADEDFIRI 220
DB 177 TIDSWILMSLTGCHGVHCTDVTNASTRLFNHSLSEMDKELEFFGIRP-----MEI 229
OY 221 GHHIVSPGTPCGNGLTAAAEMLLPCTPVAVGLIDAHAGSIGTVGEGALNNLAVF 280
DB 230 LPNVRSSSEIYG-----LKKAGALEGVPISGCLGDSALVAGOMCFODQAKN---TY 279
OY 281 GTSSCTMAST-----TSPSFVPGVWGYYSAMPGLMLVGGGSAANAIDQL 328
DB 280 GTGCGFLCNGHKCVSEHGLLTVAVKLGDKPYVAL-----EGSAVAGAVIRWL 332
OY 329 LDHFAVEAREMAQRYNOPLVWLADRILEKTAQPSDAVALAKGLHVPEFLGNRAPFA 388
DB 333 RDNIGIISSEE-----IEKLAK---EVGTSGCYFVAFSLYAPY 372
OY 389 DPHARAVICGLMERDIDLNLAYIAGLCIGIGLQILDQAQO--GVSKNIYISGAG 447
DB 373 EPSARKGICGL---TOFTNCHIAFALEAVCFQTRILDMNRDCGIRPLSHLDVDCMT 429
OY 448 OHPLVROILLDTGIGPIYITQCCPVLGSAIIGAVAGNI 487
DB 430 SKKILMOLOADILYIPVKKSPMETALGAMAAGAEV 469

RESULT 14
YDEV_ECOLI
ID YDEV_ECOLI STANDARD; PRT; 530 AA.
AC P77432; 099894;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical sugar kinase ydev.
GN YDEV OR B1511.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatter F.R., Plunkett G.III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Sequence 277,1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampa G., Seki Y., Sivasubaram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 182-495 FROM N.A.
RX MEDLINE-96243037; PubMed-8649811;
RA Das R., Reddy E.P., Chatterjee D., Andrews D.W.;
RT "Identification of a novel Bcl-2 related gene, BRAG-1, in human
RT glioma.";
RL OncoGene 12:947-951(1996).
CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCONOKINASE /
CC GLYCEROKINASE / XYLOKINASE FAMILY.
CC -1- CAUTION: WAS THOUGHT BY REF. 3 TO BE A HUMAN SEQUENCE AND WAS
CC CALLED BY THEM BRAG1 (BRAIN-RELATED APOPTOSIS GENE) (BRAG-1)
CC WITH A ROLE IN APOPTOSIS. THE DNA SEQUENCE OF THE REGION THEY
CC SEQUENCED IS MORE THAN 99% IDENTICAL TO THAT OF THIS E. COLI
CC GENE. FURTHERMORE THEY CLAIM "EXTENSIVE SIMILARITY TO THE
CC BCL-2 FAMILY OF GENES." SUCH A SIMILARITY IS NOT SIGNIFICANT
CC AND THIS PROTEIN IS MUCH MORE LIKELY TO BE A SUGAR KINASE.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; AE000248; AAC74584.1; -
EMBL; D90793; BAA15191.1; -
EMBL; D90794; BAA15198.1; -
EMBL; S82185; AAC17184.1; -
DR EcGene; EG3804; ydev.
DR InterPro; IPRO00577; FGGY_kln.
DR Pfam; PF02782; FGGY_C; 1.
DR Pfam; PF02782; FGGY_C; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; FALSE_NEG.
DR PROSITE; PS00445; FGGY_KINASES_2; FALSE_NEG.
KM Hypothetical protein; Transferase; Kinase; Complete proteome.
FT CONFLICT 490 495 PDPEKH -> TRPGA (IN REF. 2).
SQ SEQUENCE 530 AA; 57544 MW; C8C3B1E7C8982063 CRC64;

Query Match 11.5%; Score 323; DB 1; Length 530;
Best Local Similarity 24.4%; Pred. No. 5.1e-16;
Matches 142; Conservative 74; Mismatches 215; Indels 152; Gaps 26;

OY 7 VIGDVGGSVRAGIFDNLGSLSHATEK---ITTRRSGSVESQSIQAVCSTRN 63
DB 14 LMALDAGTGSIRAVIFPLEGNGIAVGAQEWNRHLAVDPVGS-MEFLNKKMQLACECMRQ 72
OY 64 ALTLADVCAQSVAGIGFDATCS---LVVLDKNGDPLFVSPEGDAKQNIYV---MDHR 115

Db 73 ALHNAIAGIEYIAAV---SACSMREGIYLYNNAGAP-----IMACANVAD 115
 OY 116 ATEOARINATHHPVLN---YVGKISPEMERPKILMLKENPEIERAGOFEDLADFL 171
 Db 116 AAREVELKELHNNTENEYRATGOTLALSAIPRLIMLHNSDSIRQASTTMTISDWL 175
 OY 172 TWARTDGLARSVCTVCKMTWLAHENRMPDEYRTIGLAELEADEP-----IRIGHNI 224
 Db 176 AYMLSGELA-----VPSNMGTTGLDLTLTRDMKPALDMAGLRADI 217
 OY 225 VSGTCCGN--GLTAQAAEM--GLPSTPAVCLIDAHAG--GICIV-----GVEGA-- 272
 Db 218 LSPVKETGLLGVSSQAELGCKAGTPVVGCGVQLCCGLGVRRPAQTAVLGCTFW 277
 OY 273 --LNNLAVYFGTSSCTMASTSPFVGWGPYYASAVPGLMVEGQSAAGAIDQLD 330
 Db 278 QOVYNLA-----APTTDEMANRV--NPH---VIRKVAEESISFTGLTMRFRD 323
 OY 331 FHRVAVE-----AREMAQRVNQPLPVM-----LADRLLEKTAQPSDAVALAK 372
 Db 324 AFCAEKEKLAERLIGIDTYLLEEMASRV--PFGSWGVPPIFSDRMFRKT----- 370
 OY 373 GLHVEPELGNRAPRAPHARAVICGLGMRDLDNLALY-----IAGLCIGYGLR 424
 Db 371 WYHAAPSEFI-----NLSIDPRCKNATLFRALENNAIVSAC---NLQ 410
 OY 425 QILDADTAGCVSKNIVISGAGOHPLVROLLADTCIGIPVITTOCCPEVLGSAIIGAVA 484
 Db 411 QIAD---FSNIHPSLIVFAGGSKGKMSQLADVSGLPNIPVVEANTALGALAGVG 467
 OY 485 GNIAPSVGEAMQOFTVHKYYP---OERYQSLHHRREYAK 523
 Db 468 AGIFSSMAETGERLVMERTHTPDREKHELUDOSRDKWQAVYQ 510

RESULT 15

XYLB_BACSU
 ID XYLB_BACSU STANDARD: PRT: 499 AA.
 AC P39211:
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Xylulose kinase (EC 2.7.1.17) (Xylulokinase).
 GN XYLB.
 OS Bacillus subtilis.
 OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxId=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Borchert S., Klein C., Piksa B., Hammelmann M., Entian K.D.:
 RT "Sequencing of a 26 kb region of the Bacillus subtilis genome
 RL downstream of spoVJ".
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Betero M.G., Bessieres P., Bolojin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
 RA Brouillet S., Bruschier C.V., Caldwell B., Capiano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
 RA Dentoz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Gim S.T., Glaser P., Goffeau A., Gollighly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joriss B., Karamata D., Kasahara Y., Klaere-Blanchard N., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Portollik S., Prescott A.M.,
 RA Prescann E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scotone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccani E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vanier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weizenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
 RA "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=85297769; PubMed=2994009;
 RA Wilhelm M., Hollenberg C.P.:
 RT "Nucleotide sequence of the Bacillus subtilis xylulose isomerase gene:
 RT extensive homology between the Bacillus and Escherichia coli
 RT enzyme".
 RL Nucleic Acids Res. 13:5717-5722(1985).
 CC -1- CATALYTIC ACTIVITY: ATP + D-xylulose 5-
 CC phosphate.
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
 CC GLYCEKINASE / XYLULOKINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 066480; AAB41094.1; -
 CC DR EMBL: 299113; CAB13645.1; -
 CC EMBL: X02795; CAA26565.1; -
 CC DR Subtilist; BG10807; xylB.
 CC DR Interpro: IPR000577; FGGY_kin.
 CC DR Pfam: PF02782; FGGY_C_1.
 CC DR TIGRFAMs: TIGR01312; xylB_1.
 CC DR PROSITE: PS00445; FGGY_KINASES_2; 1.
 CC DR PROSITE: PS00933; FGGY_KINASES_1; 1.
 CC KW transferase; kinase; xylulose metabolism; Complete proteome.
 CC SEQUENCE 499 AA; 55417 MW; 0772B7E6701E566E CRC64;

Query Match 11.5%; Score 322.5; DB 1; Length 499;
 Best Local Similarity 23.6%; Pred. No. 5,1e-16;
 Matches 130; Conservative 93; Mismatches 241; Indels 87; Gaps 20;
 OY 5 KTVIGDVSGSVRAGIFDNGSLSHATEKITTTTRSGSRVBSQSEIWAQVASCIRNA 64
 Db 2 KYVIGDILGTSAAKTLIVNQNGKVCATSKRYPLIOEKAGYSRONPDMWQOTTEALAE 61
 OY 65 LTLADVCAGSVAGIGDPAIC-SLVVLDKNGDPLVPSSEGAKNIITVMHRTBOAERI 123
 Db 62 VSIISNQAOKIDIDISYSGNHGLVLLDDODROVL-----RNAIIMNTTRTPCIRN 112
 OY 124 N---ATHNPLRVVVGKISPEMERPKILMLKENPEIERAGOFPLADLPLRATRGDGLA 180
 Db 113 TEKFGDH--LDDITKNRVLEGLPLPKMLWKHNEPELFKTAFLPKDVRRRMGVNH 170
 OY 181 RSVCTVCKMTWLAHENR--WDDPYRTIGLAELEADEPFRIGHNIIVSP--GTGCGNGL 235
 Db 171 TEVSDAA--GTLHLITRKEMSNIDICQIGIS--AD-----ICPLVESHDCVGL 217
 OY 236 TAAAAEMGLLPTPAVAVGLIDAHAGIGTVGEGCALNNLAVVFGSSCTMASTSPS- 294
 Db 218 LPHVAAKTGLETKYVAGAGADNACAGIAGILSSG--KTLCSI-GTSGVILSYEEKER 274

```

QY 295 -----FVPGVWGPYYSANVPGLMLVEGGQSAAGAIDQLDHPHVAEEAREMAQRVN 346
Db 275 DFKGKVHFFNHGKKDSFYTM-----GVTLAGYSLDMF-----KRTF 311
QY 347 QPLPWLADRILEKTAQPSDAVAL-AKGLHVPEFLGNRAPEADPHARAVICGLMERDL 405
Db 312 AP-----NESFEQLLOGVEAIPIGANGLYTPYLVGERTPHADSSIRGSLIGMDGAHNR 365
QY 406 DNLALYIAGLCIGYGLRQILDQAQTAGVYSKNIVISGAGQHPLVROILADTCGIPVI 465
Db 366 KHFLR--AIMEGITFSLHESIELFREAGKSVHTVVSIGGAKNDTWMQADIFNTRVI 422
QY 466 TTQCCPEVLLGSATLGAVAGNIAPSVGEAMQCFTHVDKYYTPQ---ERYQSLHRRREA 521
Db 423 KLENEGPPANGAAMLAAGSGMFESLECAEQFIREAAAFYPKAQNVQKYKTL---FDL 478
QY 522 YKOL-OHTAKL 531
Db 479 YKNITYHTKDL 489

```

Search completed: March 13, 2003, 16:53:44
 Job time : 12.6093 secs

QY 121 ERINATHHVNLVYVGKISPEMETPKILMLKENNPEIYERAGOFFDLATLWRATGDLA 180
 DB 121 ERINATHHVNLVYVGKISPEMETPKILMLKENNPEIYERAGOFFDLATLWRATGDLA 180
 QY 181 RSVCYVCKMTWLAHNRMPDVFRTIGLAELADEDFIRIGHIYVSPGCGNLTAQA 240
 DB 181 RSVCYVCKMTWLAHNRMPDVFRTIGLAELADEDFIRIGHIYVSPGCGNLTAQA 240
 QY 241 AEMGLPCTPVAVGLIDAHAGIGTVGVEGALNNLAIVFEGTSSCTMASTSPSFVPGVW 300
 DB 241 AEMGLPCTPVAVGLIDAHAGIGTVGVEGALNNLAIVFEGTSSCTMASTSPSFVPGVW 300
 QY 301 GPYSANVPGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEK 360
 DB 301 GPYSANVPGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEK 360
 QY 361 TAQPSDVAVALAKGLHVYPEFLGNRAPFADPHARAVICGLGERDLNLAITYGLGIG 420
 DB 361 TAQPSDVAVALAKGLHVYPEFLGNRAPFADPHARAVICGLGERDLNLAITYGLGIG 420
 QY 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLGSAIL 480
 DB 421 YGLRQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLGSAIL 480
 QY 481 GAVAGNTAPSGEAMQOFTHVDKYYPQERYQSLHRRREAYKOLQHTAKLLRD 534
 DB 481 GAVAGNTAPSGEAMQOFTHVDKYYPQERYQSLHRRREAYKOLQHTAKLLRD 534

RESULT 2

ID 052716 PRELIMINARY; PRT; 535 AA.
 AC 052716;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Ribitol kinase.
 GN RBRK.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID-573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1033-5P14;
 RX MEDLINE-97464425; Pubmed-9324246;
 RA Heuel H., Turgut S., Schmid K., Lengeler J.W.;
 RT "Substrate recognition domains as revealed by active hybrids between
 the D-arabinol and ribitol transporters from Klebsiella
 pneumoniae";
 RT Microbiology 144:1631-1639(1998).
 RN EMBL: AF045244; AAC26495.1;
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY_1.
 DR Pfam: PF02782; FGGY_C_1.
 DR TIGRFAMs: TIGR01315; 5C_CHO_kinase; 1.
 KW kinase.
 SO SEQUENCE 535 AA; 56795 MW; A5C81E4CC2F2B7F9 CRC64;

Query Match 72.3%; Score 2031; DB 2; Length 535;
 Best Local Similarity 72.3%; Pred. No. 4,7e-133;
 Matches 384; Conservative 49; Mismatches 98; Indels 0; Gaps 0;

QY 4 TKYIVGVDSGSVRACIFDLNGLSLSHATEKITTTTRSGSRVDSQSOEIWQAVCSICRN 63

DB 5 TQNIIGVDVSGSVRAGFNLRCGLAHATREITLFRBAGKRVQSSREIWOAVCYCIKT 64
 QY 64 ALTLADVCAOVAGIGFADATSLVYLDKNGDPLVSPGDKONITWMDHRAEQERI 123
 DB 65 AVANAAGVSPSSIAIGFADATSLVYLDKNGDPLVSPGDKONITWMDHRAEQERI 124
 QY 124 NATHHPVNLVYVGKISPEMETPKILMLKENNPEIYERAGOFFDLATLWRATGDARSV 183
 DB 124 NATHHPVNLVYVGKISPEMETPKILMLKENNPEIYERAGOFFDLATLWRATGDARSV 184
 QY 184 CTYVCKMTWLAHNRMPDVFRTIGLAELADEDFIRIGHIYVSPGCGNLTAQA 243
 DB 184 CTYVCKMTWLAHNRMPDVFRTIGLAELADEDFIRIGHIYVSPGCGNLTAQA 244
 QY 244 GLLTGPVAVGLIDAHAGIGTVGVEGALNNLAIVFEGTSSCTMASTSPSFVPGVW 303
 DB 245 GLLTGPVAVGLIDAHAGIGTVGVEGALNNLAIVFEGTSSCTMASTSPSFVPGVW 304
 QY 304 YSAMPVGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEK 363
 DB 305 YSAMPVGLMVEGGGSAAGAIIDQLDFHPAVEAREMAQRVNOPLPVWLADRIIEK 364
 QY 364 PSDAVVALAKGLHVYPEFLGNRAPFADPHARAVICGLGERDLNLAITYGLGIG 423
 DB 365 PSEAVTLAGLHVYPEFLGNRAPFADPHARAVICGLGERDLNLAITYGLGIG 424
 QY 424 RQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLGSAIL 483
 DB 425 RQIIDAQTAQGVSKNIYISGAGOHPLVROLADTCGIPYITTOCCPEVLGSAIL 484
 QY 484 AGNTAPSGEAMQOFTHVDKYYPQERYQSLHRRREAYKOLQHTAKLLRD 534
 DB 485 AGNTAPSGEAMQOFTHVDKYYPQERYQSLHRRREAYKOLQHTAKLLRD 535

RESULT 3

ID 0807X5 PRELIMINARY; PRT; 536 AA.
 AC 0807X5;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Ribitol kinase.
 GN AYU4324 OR AGR_L1075.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID-176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-21608550; Pubmed-11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kiteajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,
 RA Chapman P., Clendenning J., Decherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nestor E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58";
 RT Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-21608551; Pubmed-11743194;
 RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
 RA Gourollo B., Goldman B.S., Gao Y., Askenazi M., Halling C., Mullin L.,
 RA Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Wolian C., Allinger M., Doughty J., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Seac C., Strub G.,

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
 Rago G.E., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Arlitt J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Berens P.V., Bernier B.P., Bhandari D., Boltschakov S.,
 Botkina D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jostali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Mekulov S., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Zilber R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003480; AAF47823.1;
 DR FlyBase: FBgn003484; CG11594.
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF00370; FGGY_1.
 DR TIGRfams: TIGR01315; 5C_CHO_kinase; 1.
 DR TIGRfams: TIGR01315; 5C_CHO_kinase; 1.
 SQ SEQUENCE 548 AA; 59641 MW; D934CABD72A0E6B CRC64;

Query Match 35.8%; Score 1005.5; DB 5; Length 548;
 Best Local Similarity 39.4%; Pred. No. 9.6e-62;
 Matches 215; Conservative 83; Mismatches 216; Indels 31; Gaps 6;

DB 186 CKMNYDAANGSMNKEFLKQADLEELTONNFEKLSGVOPRGRTVYKGTAKAAGELGISA 245
 QY 248 GTPVAVGLIADHAGIGTVG-----VEGALNNLAVYVGTSCSTMASTSPSFV 296
 DB 246 GTVSTSLIDHAGALGFCGRSKESKGCADYOG-----KMLLIGTSCCHMSTRKACFA 301
 QY 297 PGWGPYYSAMVPGMLWVEGSGSAGAAIDQLDFHPAVEARE-----MAQRVNOPL 349
 DB 302 QGWGCPYODALIPGFLNEGGOSIAGHLDHVLKSHESYAEALKSQLGEDKFIYOHNLNKL 361
 QY 350 PVLADRLTEATAPSDVVALAKGLHVPPELGRNAPRADPHARAVTCGEMEDDNL 409
 DB 362 PELAAARL-----SOVGLTQVHWVPDHNKNSPIADPTLKGVTYGLDMTGTESLA 415
 QY 410 ALVYAGLCIGYGRLOLDAQTAOG-VVSKNIVISGAGHPVROLADTCGIPVITQ 468
 DB 416 IKYLAFLVQALVGRHIEENLYOGRAPFOFLPCGGLAKKPLVYOCADICNLPALIPD 475
 QY 469 CCEPVLGSAITLGAVAGNITVSGEAMQOFTYVDRKYYRPOERYSLHRRYEAKYQLOHT 528
 DB 476 EQEWLVGAALGAAGASGHPDLSASKSMGTQOLVKNPNETLEFHNRYKVFLOLLEN 535
 QY 529 AKLR 533
 DB 536 QROTR 540

RESULT 10
 0807J6 PRELIMINARY; PRT; 525 AA.
 AC 0807J6;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DE Ribitol kinase.
 GN AT04453 OR AGR L.826.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D., Sr.,
 Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 Kuyavian T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 Raymond C., Rouse G., Saepthimachak C., Wu Z., Romero P., Gordon D.,
 Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gatliff S., Miller N., Blanchard M.,
 Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 Flanagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,
 Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58";
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009374; AAL45247.1;
 DR EMBL: AE008240; AAR88987.1;
 KM Kinase; Complete proteome.
 SQ SEQUENCE 525 AA; 56072 MW; 0C96C97306BED7C0 CRC64;

Query Match 34.5% Score 971 DB 16 Length 525;
Best Local Similarity 40.0% Pred. No. 2.3e-59
Matches 216; Conservative 91; Mismatches 187; Indels 46; Gaps 13;

```

QY      5 KTVIGVDVSGSSVRAGIDPLDNGSLLSHAATEKITTTTTRRSGRSREVOSSOEIWOAVCSIRNA 64
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      4 ONLVADVDTGSASRAGICTDPAGRLLRARSIPHLTMQRPRENNAEHNDSTDMNNVAICAVKAA 63
QY      65 LTLADVCASVAGIGFEDATCSLVLDKNQDLPVSPEGDAKONITVMMDHRTAEQAERIN 124
Db      64 LADAGVLTPQSIAAIEFDATCSLVIFDERGEPVSVSTDDREDFTIVMLDHRAIGEADELUT 123
QY      125 ATTHPVLNYVGSKISPEHEETPKILMKENMPETIERACGFPLDALPTVRAIGDLARSC 184
Db      124 ASGHRVLDVFAGNSVSPQMOPKLMTLKTHPMSWSRMSFPALDFLTWKATGSAQRNSC 183
QY      185 TVTCKWTLAHEN-RWDDPYFTTCLAELEADDFTIRIG-NHIIVSGTPRCNGTLTAQAAA 241
Db      184 TOTAKMNTLADEHPNQWDYLAFAGLDLIKE-----RAGLPETTVMPGAI-G-PLSEPAFA 238
QY      242 EMGLTGPRVNAVGLIDAHAAGIGITGVVEGAL-----NNLAVVFSTSSCTMASTTSPSFV 296
Db      229 ELGIDTGGCVAAGMIDAVAGAALGAL--GCCLAEDVGNKVALLAGTSGLVAMSNQP--M 293
QY      297 PG--VMGPYYSAMPDGLNLVEGSGSAGAADIQLDFHNPAREARMARVNOLPFWLDA 354
Db      294 PGRSLMGTYMOAVLTPGHNLVEGSGSATGALLDHIIVRNAAGSEP-----DTALHARI 346
QY      355 DRIILEKTQPSDAVALANGIHVPFFLGNRAPFADPHARAIVCGISEMERDENLLALUYTA 414
Db      347 ARVTE--LRELEGEFARLRLHVLPFHNRSLAPLARHAGVVSGLTDFSISLCRLMYR 404
QY      415 GLCGIGYELROILOQTQOGVSVKNIVISGGAGRPVROLIALDRPGIVITTOCECPYL 474
Db      405 TATAALGARHYLDLMERFGVAYETLHTYGCHVKRPMLELADYTKGIIVPATADAVL 464
QY      475 LGSAILGAIVAGNIAPSV---GEAMQO-----FTYHDKYXYPOERYOSLNHRYE 520
Db      465 LGTAMTATYAGGVHASLAAAGAAAMPRGNAEISGNPALAAHYERYD---RRFLAMYRHQE 521

```

RESULT 11
Q96C11 PRELIMINARY; PTR: 439 AA.
AC Q96C11:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 47.7 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKIN;
RA Strausberg R.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC014947; AAL14947.1; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF02782; FGGY_C.1.
DR TIGRFAMS: TIGR01315; 5C_CHO_kinase; 1.
KM Hypothetical protein.
SQ SEQUENCE 439 AA; 47706 MW; 60F2354474EBB3E8 CRC64;

```

OY 169 DEFMRARAGDLARVSCVYTCWKKWMLAHENRMDPRDPEFTIGLAELADEDDFRIGHNIHSFG 228
Db 61 DFLSKRAKAGVYARSLCSLVCKMWTSA-EKGDSDSFMMKIGLEDDVADANYSKIGNQVLPBG 119
OY 229 TPCCNGGLTAAQAAMGMLPGTPRVAVGLIDAHAGIGTVG--VEBGAL-----NNLAY 278
Db 120 ASLNGGLTPPEARLDGLLPGIAVAASLIDAHAGSLGYIADYVRHGLICSGQVYTSRLAV 179
OY 279 VEGTSSCTMASTSTSPSPFVPGVGPVYYSAMVRGLMLVEGGSAAAGALIDQLLDFHPAYEEA 338
Db 180 ICGTSSCHMGISKRPFIKPVGCVGSGYFESAMPGEFWLNEGSGSVTGKLLIDHWQCHAAFPRL 239
OY 339 REMAQRVNQRPLVYLADI--LEKPAOSSDAVALAKGLHVPELIGNAPRADPHARVTC 397
Db 240 QVKATARCQSYATVLSHSLDITKRAOP--VGFLTVLDLHWPDDEGNSPLADLTGLGMVT 297
OY 398 GLGMRDLCIDNLALAYIAGLIGCYGLRQIILDAOTAGCVSNNTVYISGAGQNHPLVRQILA 457
Db 298 GLKLSQDLDLALILYLANVQAIALGTFTIIEAMAAHSHISTLEFLCGSLSNPLFYQMHA 357
OY 458 DTGCIPIVITTCCEBPVLGSAIILCAVAGNIAPSVGEAMQOFTYHDKYUYUQERYQSLNHR 517
Db 358 DITGMPVVLISQEVESVILGAVALTGACASGDFAVSQEAMAMKSKVGKVVFPRLDDKKYYDK 417
OY 518 RYEAV-KOLOHTAKLL 532
Db 418 KYQVFLKLVENHQEYL 433

```

```

RESULT 12
096UC2
ID 096UC2 PRELIMINARY; PRT; 598 AA.
AC 096UC2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Conserved hypothetical protein.
GN 61d6.030.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
NCBI_taxid=5141;
RN (1)
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nykattura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA German Neurospora genome project.
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AL513408; CAD1134.1; -.
interPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY; 1.
DR TRGFAMS: TRGR01315; 5C_CHO_kinase; 1.
DR Hypothetical protein.
KW SEQUENCE 598 AA; 64896 MW; 9A91804D82709195 CRC64;

```

RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC014947; AAI14947.1; -
 DR InterPro: IPR000577; FGGY_kin.
 DR Pfam: PF02782; FGGY_C; 1.
 DR TIGRFAMs: TIGR01315; 5C_CHO_kinase; 1.
 KW Hypothetical protein.
 SO SEQUENCE 439 AA; 47706 MW; 60F2354474EBE3E8 CRC64;
 Query Match 33.4%; Score 940; DB 4; Length 439;
 Best Local Similarity 44.5%; Pred. No. 2,6e-57;
 Matches 194; Conservative 78; Mismatches 140; Indels 16; Gaps 7.
 QY 110 VMMDHRAEQAEIRINATHHPVLNVYGGKISPEMETPKILMKENPEIYYERAGOFFDIA 168
 :||||| | ||| | ||||| :||| :||||||| :||| :||| ||||
 Db 1 MMLDHRRAVSQVNRINIEHTKHSVLQYVGGVMSVEMQAPKILMKLENRLREICWDKAGHFDLP 60

[illegible]

QY 176 TGDARSVCIVCKWTL-----AHENKMDPDYFRTIGLAELADEDFIRIG-----HHI 224
D 186 TGKEIRSEFCSTCKOGKGVGVGSDSVKGMQEDFEYETIGLGLYDNTFRMGVGVNGSW 245
QY 225 VSPSTPCNGTLTAQAAEMGLLPCTPAVAVGLIDAHAGIGTVGE-----269
D 246 MSAGELVGG-GLSEAGGQGLPAGIAGSGVLDYVAGVIGVAKVGLSPDRDSDVAPN 304
QY 270 --GGALNNLAVFCTSSCTMASTSPSFVPGVWGPYSAVPGMLVYEGGSAAGAIDQ 327
D 305 DVSGAFRLAAVAGTSTCHLAMSGLPVFPGVWGPYRVLVPGVMAEGGSAIGELMKH 364
QY 328 LLDHPRAVEEREMAKOVNPLPVWLADRIE---KTAQPSDAVALAKGLHVPBEFLGNR 384
D 365 MLEHNVAVDETVKAEAGAKNIYDLNHELMKAKTAPISY-LVHFFFGYDGLMGNR 423
QY 385 APFDPHARAVYICGLGEMRDNLALYIAGLIGYGLRILDAQTAQGVSKNIVISG 444
D 424 SPIDPNNKRGATIGMSNDKSGOMALLYSTMERIALOTROIVEMNTAGHTISFMSG 483
QY 445 GAGHPPLVROILLADTCGIPVITTOCE-PVLLGSAIIGAVAGNIAPS-----VGEAMQF 498
D 484 SQCNELIMDLIATACDMPVLIPIRYVNAAYVGAAMLGAKAASAKPOSTTEPLMDINDRM 543
QY 499 THVDKYYT---POERTQSLHRRYEAVKOLQHTAKLR 533
D 544 SKPGKTWVSRGDPAEK--KLDAKYEIIFLDARTQOEYR 580

RESULT 13

ID 004585 PRELIMINARY; PRT: 715 AA.

AC 004585;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 79.2 kDa protein.
GN YDR109C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Murphy L., Shore L., Harris D.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrett B., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z48758; CA88663.1; -
DR SGD: S0002516; YDR109C.
DR InterPro: IPR000577; FCGY_kin.
DR Pfam: PF00370; FCGY_1.
DR TIGRFAMs: TIGR01315; 5C_CHO_kinase; 1.
KW Hypothetical protein.
SQ SEQUENCE 715 AA; 79160 MW; A52B2D9B4B4C9C7E CRC64;

Query Match 30.8%; Score 864.5; DB 3; Length 715;
Best Local Similarity 36.1%; Pred. No. 9.1e-52;
Matches 201; Conservative 100; Mismatches 211; Indels 45; Gaps 10;

QY 5 KTVGVGVGSGSVAGIFDUNGSLSHATEKITTRSGSVHEQSSOEIMQVSCIRNA 64
D 38 KFYGVGVGVGSGARACYIDOSGNMLSLAEKTRQQLINSFTOSSREIMAVCYCVRTV 97
QY 65 LTLADVCAQSVAGIGFATSLVLD-KNGDPLVSPF-GDAKONITVMDHRAETEQAER 122
D 98 VEESGVDPERVRIGIGFATSLVYVSATNFEIAGVDFTNNDNIIMDMHRAKETEER 157

QY 123 INATHHPVLNVGKISPEMETPKILMKENMPETIYERAGOFEDLADFLTWRAATGLARS 182
D 158 INSSGDKCKLVGCGMSVEMEIPIKMLKNLEAGIEFQDCFFDLPDYLFKAKGKENS 217
QY 183 VCTVTCWKWTL-----AHENKMDPDYFRTIGLAELADEDFIRIG-----HHIYSPGPC 231
D 218 FCSAVCKQGLFVGVGSDIGMSKEFLNSIGLSLTKNDFERLGLSKREKNFILA-EC 276
QY 232 GNGTLTAQAAEMGLLPCTPAVAVGLIDAHAGIGTVGE-----GGAL 273
D 277 ISPLDKKAAQGLGLEHCVSSGIIDAYAGVGVYAAKPESAVGLAETENYKKDFNCAL 336
QY 274 NNLAVFCTSSCTMASTSPSFVPGVWGPYSAVPGMLVYEGGSAAGAIDQLDHP 333
D 337 GRLLAVAGTSTCHLILSKNPIFVHGVGPRDYVLARGMAEAGOSQCGVLLDLITTHP 396
QY 334 AVEEREMAKOVNPLPVWLADRIEKTAPQ---SDAVALAKGLHVPBEFLGNRAPFADP 390
D 397 AFTELSHMANLAGSKFEYL-NKILETLVERKRYRVSLSLKLHFFYGDYGNNSPIADP 455
QY 391 HARAVYICGLGEMRDNLALYIAGLIGYGLRILDAQTAQGVSKNIVISGAGQHP 450
D 456 NMRACIIQSMDNSIEDLAVYLSACEFTISQOTROIIVMLKSGHEINAIEMSGQCRNS 515
QY 451 LVROILLADTCGIPVITTOCE-PVLLGSAIIGAVAGNIAPS-----VGEAMQFTHVDKYYTPOE 509
D 516 LLMRLADCTGLPIVPIRYVNAAYVGAALGAAASDE-----DYTREKRTLAGOK 567
QY 510 RYQSLHRRYEAVKOLQ 526
D 568 SSQTKTERFENDSYSSIQ 584

RESULT 14

ID 095T84 PRELIMINARY; PRT: 439 AA.

AC 095T84;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE GH12991P.
GN CG11594.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Abbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY060286; AL25325.1; -
DR FLYBase: FBgn0035484; CG11594.
DR InterPro: IPR000577; FCGY_kin.
DR Pfam: PF02782; FCGY_1.
DR TIGRFAMs: TIGR01315; 5C_CHO_kinase; 1.
SQ SEQUENCE 439 AA; 48172 MW; F26903DBB321A40E CRC64;

Query Match 27.4%; Score 769.5; DB 5; Length 439;
Best Local Similarity 37.9%; Pred. No. 1.8e-45;
Matches 167; Conservative 69; Mismatches 176; Indels 29; Gaps 5;

QY 112 MDRATEQAEIRINATHHPVLNVGKISPEMETPKILMKENMPETIYERAGOFEDLADFL 171
D 1 MDHRAEQETQIRINAFKSLILKXYGGVSLMEVYFKLMLKNLSQTFGINIRVDFLDL 60
QY 172 TWRAATGLARSVCIVTCWKWTLAHENKMDPDYFRTIGLAELADEDFIRIGHIYSPGPC 231

```
Db 61 TWRATGVDRSLCSVYCKMNYDANGSMWKEFLKQADLEBLTONNEKLSGDVQPPGRIV 120
OY 232 GNGLTQAQAAEMGLPPTPAVGLIDAHAGIGTVG-----VEGALNNLAYVF 280
Db 121 GKCLTKAKAGELGSLGTVVSTSLIDHAGALMFGRSKESGADVDG---KKAFLA 176
OY 281 GTSSCTMASTTSPSPVGVWGPYYSAMVPGMLVEGQSAAGAIDOLDLPHFAVEARE 340
Db 177 GTSTCHMSITRKACFGAGWGPYQDAIIRGYFNEGQSIAGLLDHVLSKSHSYAELKS 236
OY 341 -----MAQRVQPLPVLADRILEKTAQPSDAVALAKGLHVPPEFLGNRAPFADPHAR 393
Db 237 QLEGEDEFYOHNLKLLPELAAARGL-----SQVGCLTDVHWVPLDGNRSPADPTLR 290
OY 394 AVICGCGMERDLDNLALYTAGLCIGYGLRQLDPAOTAG--VYSKNIVISGAGOHPLV 452
Db 291 GVTTGDMTRGTESLAIKTYLAFQALAYGTRHLEINLYOYGRAPFQTLFCGGIANKPLX 350
OY 453 RQIADTCGIPYITTOCCEPVLGSAIILGAVAGNIAPSVGEAMQOFTHVDRYYPQERYQ 512
Db 351 VQCHADICNRPALIPDEQENVLYGAALGAAGHFDLSLESASKMGCTQLYKPNARETL 410
OY 513 SLHRRYEAVKOLQHTAKLR 533
Db 411 EFHRRKYKVFQLEENQRYR 431
```

RESULT 15

```
O9HA63 PRELIMINARY; PRT; 246 AA.
ID O9HA63
AC O9HA63;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CDNA FLJ12175 f1s, Clone MAMMA1000713, weakly similar to
DE L-ribulokinase (EC 2.7.1.16).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niimiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK022237; BAB13993.1; -
SQ SEQUENCE 246 AA; 26576 MW; 04D5374635B4DB01 CRC64;
```

Query Match 23.9%; Score 672; DB 4; Length 246;

Best local Similarity 52.0%; Pred. No. 4, 8e-39;

Matches 127; Conservative 38; Mismatches 67; Indels 12; Gaps 4;

```
OY 110 VMQDHRTGAEERINATHNPLVNYGKISPEMETPKIIMLKEMPEI-YERAGQFFDLA 168
Db 1 MVIDHRAVSVQVNRINETHKHSVLQYGVGMSVEMQAPKLLWKELREICWDKAGHFDDL 60
OY 169 DELTWRATGDLARSVCYVCKKWTMLAHENWDDYFTTIGLAEADDEFIRIGHIYSPG 228
Db 61 DFLSMKATGYTARSLCSLVCKWYSA-EKGMWDSFWMKMIGLEDVADNYSKIGNQVLP 119
OY 229 TPCGNGLTQAQAAEMGLPPTPAVGLIDAHAGIGTVG--VEGAL-----NNLAY 278
Db 120 ASIGNGLTPEAARDGLPGLIAVAASLIDHAGLGIVGADVGRHGLICGQPTSKLAY 179
OY 279 VFGTSSCTMASTTSPSPVGVWGPYYSAMVPGMLVEGQSAAGAIDOLDLPHFAVEEA 338
```

```
Db 180 ICGTSSCHMGISKDPIFFVPGWGPYFSAMVPGFWLNEGOSVYGLIDHNVQCHAAPEL 239
OY 339 REMA 342
Db 240 QVKA 243
```

Search completed: March 13, 2003, 16:55:12
Job time : 37.0381 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:37 ; Search time 34.7483 Seconds
(without alignments)
1625.927 Million cell updates/sec

Title: US-09-802-208B-5
Perfect score: 2211
Sequence: 1 MSRRNKQWIGLPLHLWGVI.....RVQPGFSSAPVTEKALNIS 424

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries:

Database : A.Geneseq.101002:*

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2211	100.0	424	22	AAE09781
2	1040.5	47.1	467	22	ABP38175
3	580.5	26.3	445	22	AA689873
4	174.5	7.9	414	20	AAV37671
5	168.5	7.6	457	21	AA18235
6	164	7.4	394	22	ABB47287
7	158.5	7.2	387	23	ABB53425
8	158.5	7.2	416	20	AAV02610
9	158.5	7.2	416	22	AA898023
10	158.5	7.2	416	22	AA84207

11	155	7.0	397	23	ABP28804
12	152.5	6.9	392	22	AAE05852
13	147	6.6	401	23	ABB48316
14	145	6.6	393	23	ABP40749
15	143.5	6.5	484	22	AA676735
16	143.5	6.5	1595	22	AA693189
17	142	6.4	497	22	ABB60537
18	141	6.4	352	20	AAV35284
19	141	6.4	496	23	ABB54015
20	140	6.3	447	22	AA848873
21	139	6.3	417	22	AAU34999
22	138.5	6.3	449	22	AA692801
23	138	6.2	388	18	AAW26740
24	138	6.2	454	22	ABB52906
25	137	6.2	382	23	ABB47392
26	137	6.2	405	23	ABB53977
27	136.5	6.2	425	22	AAU34874
28	136.5	6.2	425	22	AA698958
29	136	6.2	492	23	ABP40002
30	133.5	6.0	490	23	ABP30110
31	132.5	6.0	378	22	AA692975
32	132.5	6.0	402	22	AA676763
33	132.5	6.0	412	22	AA678973
34	132.5	6.0	423	23	ABP40365
35	132	6.0	430	20	AAW88344
36	131	5.9	433	23	ABP27912
37	131	5.9	466	22	ABB52555
38	130.5	5.9	467	23	ABB53687
39	130.5	5.9	542	22	ABB66844
40	130	5.9	518	22	AA635474
41	130	5.9	701	22	AA692442
42	129.5	5.9	453	22	AA693123
43	129	5.8	400	23	ABP27911
44	129	5.8	400	23	ABP29844
45	129	5.8	401	21	AAV91291

ALIGNMENTS

RESULT 1
AAE09781
ID AAE09781 standard; Protein; 424 AA.
XX
AC AAE09781;
XX
DT 29-NOV-2001 (first entry)
XX
DE Escherichia coli strain C rfl operon encoding ribitol transporter.
XX
KW Positive selection system; metabolise; arabinol; ribitol; mannitol;
KW transgenic cell; marker gene; ribitol dehydrogenase; ribitol kinase;
KW ribitol transporter; rfl operon.
XX
OS Escherichia coli C.
XX
PN WO200166779-A2.
XX
PD 13-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US07474.
XX
PR 08-MAR-2000; 2000US-0188291.
PR 15-AUG-2000; 2000US-0255595.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Parrott W, Lafayette P, Kane P;
XX
DR WPI; 2001-565596/63.
DR N-PSDB; AADI6811.
XX
PT Positively selecting transformed cells comprising selectable marker

Streptococcus poly
Pseudomonas stutzeri
Listeria monocytog
Staphylococcus epi
Corynebacterium g1
C glutamicum prote
Drosophila melanog
Chlamydia pneumonia
Lactococcus lactis
Lactococcus lactis
Enterococcus faecia
C glutamicum prote
Staphylococcus car
Escherichia coli p
Listeria monocytog
Lactococcus lactis
E. coli cellular p
E. coli growth and
Staphylococcus epi
Streptococcus poly
C glutamicum prote
Corynebacterium g1
C. glutamicum SRT
Staphylococcus epi
Salmonella enteric
Streptococcus poly
Escherichia coli p
Lactococcus lactis
Drosophila melanog
B cepacia multidru
C glutamicum prote
C glutamicum prote
Streptococcus poly
Streptococcus poly
Group B Streptococ

PT gene and desired gene, from a cell population by using marker compounds
PT e.g., arabinotol, ribitol which confer selective advantage on transformed
PT cells
XX
XX
PS Claim 18; Page 37; 37pp; English.
CC The present invention relates to a positive selection system that
CC involves conferring to transferred cells the ability to metabolise
CC arabinotol, ribitol and/or mannitol. The positive selection method is
CC used in positively selecting transgenic cells from a population of cells
CC using the positive selection method, the presence of the gene of
CC interest in the genetically transformed cells may be determined without
CC the disadvantages associated with traditional negative selection
CC systems. Positive selection of the transformed cells is achieved without
CC directly damaging the neighbouring non-transformed cells. The
CC transformed cells may be identified by simple visual means without the
CC use of a separate assay to determine the presence of a marker gene. This
CC technique also avoids the release of antibiotics or other dangerous
CC genes into the environment. The present sequence is Escherichia coli
CC strain C ribitol transporter encoded by ribitol operon (rtl operon). The
CC operon also encodes ribitol dehydrogenase and ribitol kinase.
XX
XX
SQ Sequence 424 AA:
Query Match 100.0%; Score 2211; DB 22; Length 424;
Best Local Similarity 100.0%; Pred. No. 3.3e-227;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRRNKMWLPRLHNGYIAIAFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLA 60
DB 1 MSRRNKMWLPRLHNGYIAIAFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLA 60
QY 61 AALSAMWSGVVAEITTPQKTMILGFLVLCVFNHFLVFLVGLQANVGLILFFYIGRLAYP 120
DB 61 AALSAMWSGVVAEITTPQKTMILGFLVLCVFNHFLVFLVGLQANVGLILFFYIGRLAYP 120
QY 121 LFLYSFIVIIHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALA 180
DB 121 LFLYSFIVIIHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALA 180
QY 181 FCFPGGYIAMSIRHVTTPGHHNLTPREKPAELSRVATLLTYRNIFLSIVHINTLS 240
DB 181 FCFPGGYIAMSIRHVTTPGHHNLTPREKPAELSRVATLLTYRNIFLSIVHINTLS 240
QY 241 LFGFAVIMPMWVDELGFTTSEMLQVMAAFETTFISNIFEMGIVAEKMGKMRVRWFGCL 300
DB 241 LFGFAVIMPMWVDELGFTTSEMLQVMAAFETTFISNIFEMGIVAEKMGKMRVRWFGCL 300
QY 301 GMAASSLAFFYIMPOYFGHNWMMAMIPALIGTFVAAEFVPMMAVFPALPEPKHGAISYVN 360
DB 301 GMAASSLAFFYIMPOYFGHNWMMAMIPALIGTFVAAEFVPMMAVFPALPEPKHGAISYVN 360
QY 361 LSAGMSNFLPAIVVLLPWFSTIGVIAVATLALFVLCARFVPOPGSSAPVTEKA 420
DB 361 LSAGMSNFLPAIVVLLPWFSTIGVIAVATLALFVLCARFVPOPGSSAPVTEKA 420
QY 421 LNIS 424
DB 421 LNIS 424
RESULT 2
ABP38175
ID ABP38175 standard; Protein; 467 AA.
XX
XX
AC ABP38175;
XX
XX
DT 24-JUL-2002 (first entry)
XX
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3020.
XX
XX
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
antibacterial; gene therapy.

XX
OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 9805-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI, 2002-381255/41.
XX
XX N-PSDB; ABN90720.
XX
XX Novel isolated nucleic acid encoding a staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
PS Disclosure; SEQ ID 3020; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
XX
SQ Sequence 467 AA:
Query Match 47.1%; Score 1040.5; DB 23; Length 467;
Best Local Similarity 48.6%; Pred. No. 3.3e-102;
Matches 195; Conservative 76; Mismatches 129; Indels 1; Gaps 1;
QY 10 GLPLHLIMGYIAIAFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLAALISAVSG 69
DB 55 GLPLSLIMGFVAIAIAFMGDDGFEFLAFSLHYIKSLGFTPAEASFATLYGLAALISAVSG 114
QY 70 VVAETTPQKTMILGFLVLCVFNHFLVFLVGLQANVGLILFFYIGRLAYPLFYSTIYV 129
DB 115 VLAETTPQKTMILGFLVLCVFNHFLVFLVGLQANVGLILFFYIGRLAYPLFYSTIYV 174
QY 130 IHNVRSESSSALGMYAVSVGIGVAGSYTPSTPIMGEMTLMALAFCPAGGVIA 189
DB 175 IAOVTPGARMAASANGWMSMCIGIGLGNWIPBLSIRGIFINTLWGVFWVAVALMI 234
QY 190 MISLRHVKTPGHHNLTPREKPAELSRVATLLTYRNIFLSIVHINTLSLFGFAVIMP 249
DB 235 MYLWKEGAGKRPDAVTLVERLKLSSGVIIARRRGMFLVIRIICNLSLFGFLPILP 294
QY 250 MMFVD-ELGFTTSEMLQVMAAFETTFISNIFEMGIVAEKMGKMRVRWFGCL 308
DB 295 LYLTSVEGFSMEQMLHMGVMEFVASTVNIWQIDRLGMLQOMWFGCIGALSLLA 354
QY 309 FYVMPQYFGHNWMMAMIPALIGTFVAAEFVPMMAVFPALPEPKHGAISYVNLSAGMSNF 366
DB 355 FYVLPQYGAHFELALAAIGFISVAFPMGAVFLALAPDEKGAISAHNLAAAGSNF 414
QY 369 LAPAIIVVLLPWFSTIGVIAVATLALFVLCARFVPOPGSSAPVTEKA 409
DB 415 MGPGIATLFIATLIGKGVWVIYAGLVAAVAGVLPFFIVPOP 455
RESULT 3
AAG89873

PN WO200179257-A2.
XX 25-OCT-2001.
XX 12-APR-2001; 2001WO-US12230.
XX 14-APR-2000; 2000US-197349P.
XX (PHYT-) PHYTERA INC.
XX Davis DV, Rogers BL, White AC;
XX WPI; 2001-626526/72.
XX N-PSDB; ABA82962.
PT Determining whether a candidate nucleotide or polypeptide
PT encodes/functions as a multidrug resistance (MDR) efflux pump comprises
PT searching a database of nucleotide/polypeptide sequences for those with
PT high identity to known MDR pumps -
XX
XX PS Claim 10; Fig 28; 139pp; English.
XX
CC The invention relates to determining whether a candidate nucleotide
CC (ABA82338-ABA82971) or polypeptide (ABA47263-ABA47296) encodes/functions
CC as a multidrug resistance (MDR) efflux pump comprising, searching a
CC database for sequences high identity known MDR efflux pumps and then
CC deleting/mutating an identified region of the DNA in a bacterial cell and
CC determining whether the bacterial cell exhibits increased or decreased
CC sensitivity to an antibacterial agent The identified pumps are useful for
CC further identifying MDR efflux pumps that may be used as drug targets to
CC increase the sensitivity of cells to antibacterial agents. Cells
CC comprising the identified pumps may be used to screen for potential
CC blockers or inhibitors of MDR pump function or gene expression.
XX
XX SQ Sequence 394 AA:
Query Match 7.4%; Score 164; DB 22; Length 394;
Best Local Similarity 20.8%; Pred. No. 9.3e-09;
Matches 86; Conservative 76; Mismatches 172; Indels 80; Gaps 18;
QY 19 YTAIA-----VFMTGDFELAFISHYIKSLGFTPAEASFAFTLYGLAAASAMVSGVAEI 74
DB 19 YTAISNLEFLVFL-GVGLVLPVLPOLKEMHFSGTTMGMMISIFAIAQLITSPAGVLSDK 77
QY 75 ITPQTMILGFVLCVHFVFLFVFGIGQANGYGLILFY---GIRGLAYFLFYSLFVIL 131
DB 78 IGRKMIATGML---VFSISELDFGLAQKSG---FYISRGIGIAAALMLPSTAYEVA 130
QY 132 HNVRSNSSSALGWYAVYS---VGIGVAGSYIPSTPIPMGEMGTMLALAFCPAGCV 187
DB 131 DMTTISERKAMGIVSAISGFIIGPGVG-----FLAYGIRAPFPFAAFIATIGFI 184
QY 188 IAMISIRHVKTPGHMHNLTPREKF--AELSRATVLLYTRNIFLSS--IVRIINTLSLFG 243
DB 185 LRLTVLKE-----PEKRIILAAVEAKKSGFMDILRNPMFTSLFYIIILISSGLQA 233
QY 244 FAVIMPMVDELGFTTSE-WIQWMAAFFTTIFSNIFGIVAEKKGMKRVIRMECLGM 302
DB 234 FESISIMATINFGFTSETAIVTVSGILALCOLFFPDALVOKIGEGLLQIOL----- 287
QY 303 AASSLAIFYMPQFGHNYMMAMP-----AIALGTFVAAFVMAVPALE---PKHKG 353
DB 288 -----TFPASAIFIAVIAFTKNNLVVFSFEI-VPLAFDLFRAAVTTYSKHNKG 335
QY 354 AAIISVYN---LSAGMSNFLAPAIAYVL-----LPMESTIGVYI--AVTALYL 395
DB 336 DOOGTINGLNSTFTSPGNILGPMAGALFDINHFEPYVAVSAILLCTGSLFTL 389
RESULT 7
ABBS3425
ID ABB53425 standard; Protein; 387 AA.
XX

AC ABB53425;
XX 16-MAY-2002 (first entry)
DE Lactococcus lactis protein b1t.
XX Lactococcus lactis IL1403.
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX Lactococcus lactis IL1403.
XX FR2807446-A1.
XX 12-OCT-2001.
XX 11-APR-2000; 2000FR-0004630.
XX 11-APR-2000; 2000FR-0004630.
XX (INRG) INRA INSTR NAT RECH AGRONOMIQUE.
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX New nucleotide sequence useful in the identification or Lactococcus
XX Lactis and related species -
XX
XX PS Claim 6; SEQ ID NO 127; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX SQ Sequence 387 AA:
Query Match 7.2%; Score 158.5; DB 23; Length 387;
Best Local Similarity 21.0%; Pred. No. 3.5e-08;
Matches 89; Conservative 78; Mismatches 180; Indels 77; Gaps 19;
QY 3 RNNKOWILGPLHLIMGYIAAVFMGDFELAFISHYIKSLGFTPAEASFAFTLYGLAAA 62
DB 2 KKKKSMNNLAISNLF-----LVFL-GAGLVIPVLPOLKEMHFSGTTMGMMISIFAIAQL 55
QY 63 LSAMVSGVAVEIITPOKTMILGFVLCVHFVFLFVFGIGQANGYGLILFY---GIRGLAY 119
DB 56 VASPAVAGALSDKIGRKRKLAIGMI---IFSSELDFGLAQKSG---FYISRALGCVAA 108
QY 120 PLFVFSYIVYIHNVRSNSSSALGWYAVYS---VGIGVAGSYIPSTPIPMGEMGTL 175
DB 109 AMIMPSTVAVADMTTIAERPRAMGLVSAISGFIIGPGV--GGFLAHGIRV-----FP 162
QY 176 WLALAFCPAGGVYAMISLRHVKTGPHMHNLTPREKFAELSRATVLLYTRNIFLSS--IV 233
DB 163 YVAAILAFILGITLITLKEPERTIESHOIEKVSFLDIL-----KNPLFGSLFII 213
QY 234 RIINTLSLFGPAVIMPMVDELGFTTSEWQWMAAFFTTIFSNIF-WGIYAEKKGMKRV 292
DB 214 IILSSGLOAFESISYISIMSGFMSSEALVTVSGILALFQFLDVAIVNKIGELG 273
QY 293 VIRMFQCLMAASSLAIFYMPQFGHNYMMAMP-----AIALGTFVAAFVMAVPA 346
DB 274 LTIOL-----TFPASAIFIAVIAFTKNNLVVFSFEI-VPLAFDLFRAAVTTYSKHNKG 315
QY 347 LE---PKHKG---AAISVYNLS--AGMSNFLAPAIAYV-----LPMESTIGVYI--AVT 391
XX

XX Telford J, Maslmani V, Margarit Ros YI, Grandi G, Fraser C;
PI Telford H;
XX WPI: 2002-352536/38.
DR N-PSDB: ABN69435.
XX
PS Claim 1: Page 3841: 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (I), nucleic acids encoding (II), ABN66044-ABN71516 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 397 AA:

Query Match 7.0%; Score 155; DB 23; Length 397;
Best Local Similarity 23.6%; Pred. No. 8.6e-08;
Matches 100; Conservative 76; Mismatches 161; Indels 86; Gaps 24;

QY 8 WLGPLHLINGYIAIAVFMGDFELA--FLSHYIKSLGFTPAE----ASFAFTLYGLA 60
DB 11 WLG-----NFTGASFSLVMPFMAVVENLG--TPRELVEYAGLAVAAYALA 56
QY 61 AALSAMVSGVAELITPQKTM--IGFVLMCVFHLFLVFGG--QANGGLILFVGIGLA 118
DB 57 SALFAPWGLADRYGKRPMLRASFVW-----TFMGGALITPNVFWLILLLLTGVS 110
QY 119 YPLVLSFIVIIHNVSESSSALGWYVAVSVGIGV--AGSYPTSTPIPMGEMLTL 177
DB 111 AG-YVPAPATLIALSQAKESGYALG-----TLATGTGALIGLPLGLLAE--LGI 161
QY 178 ALAFCEPAGVIAIMISLR---HVKTGGMHMLTPREKFAELSRVATLLYTRN----NIFL 229
DB 162 ROVELLAVGVILFLCISLMTAVYVK-----EEFKPYRREMIPKVIILQVKSPOIMLGFV 216
QY 230 SSIVRIINTSLFGEFAVIMPMFVDELG-----FTSEMLOWAAFEFTTISNIMGI 283
DB 217 TSMITQISAOVS---APIL-SLYIRHLGQTHNLMFTSG---LVVASGFSLSSTSLGK 269
QY 284 VAEKMGMYRIYRFGCLGMAASSLAF--YVMPDYFGHNYMMAMIPALGTFVAAPV-MA 341
DB 270 LGRBFGRHRL-----LALCYSFIMYSSALAOFSFOLGVLRFYGVGVGLMPSIN 322
QY 342 AVFPALEPKKGAIAISYV--LSAGMSNFLAPALAVLLPWFSTIGVIAVATLY----LL 396
DB 323 SLTLKTLPKGIGISRVFAYNOMFSLGQVIGPFG-----SNVAVVLGYSRVFVYVSLI 375
QY 397 AFV 399
DB 376 VFV 378

RESULT 12
AAE05852
ID AAE05852 standard; Protein; 392 AA.

XX AAE05852;
AC 24-SEP-2001 (first entry)
DT XX
DE Pseudomonas stutzeri open reading frame-N (ORF-N) protein.
XX
KW Open reading frame-N; ORF-N: pyridine-2,6-bis (thiocarboxylate);
KW PDTC; environmental remediation; phytoremediation; bioaccumulation;
KW water purification; solution mining mobilisation; immobilisation;
KW detoxification; redox state modifier; metal ion reactivity;
KW carbon tetrachloride; metal.
XX
OS Pseudomonas stutzeri.
XX
PN WO200153309-A1.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001MO-US02386.
XX
PR 20-JAN-2000; 2000US-0177251.
XX
PA (IDAH-) IDAHO RES FOUND INC.
PA (PASZ/) PASZCZYNSKI A.
PA (SEBA) SEBAT J L.
XX
PI Paszczynski A, Sebat JL, Lewis TA, Crawford RL, Cortese MS;
XX
DR WPI: 2001-465361/50.
DR N-PSDB: AAD11172.
XX
PT New isolated nucleic acids, useful for producing enzymes required to
PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for
PT reducing the amount of metal or carbon tetrachloride in a substrate,
PT e.g. soil or water -
XX
PS Claim 32: Page 100-101; 172pp; English.
XX
XX The present sequence is a Pseudomonas stutzeri open reading frame-N
XX (ORF-N) protein. The Pseudomonas stutzeri genome includes ORFs that
XX encode enzymes required for biosynthesis of pyridine-2,6-bis
XX (thiocarboxylate) (PDTC). The ORFs encoding PDTC are especially useful
XX in environmental remediation methods, e.g. phytoremediation,
XX bioaccumulation, water purification, waste water purification, solution
XX mining mobilisation, immobilisation, detoxification, redox state
XX modifier or modification of metal ion reactivity. In particular, the
XX ORFs are useful for degrading carbon tetrachloride and removing metals
XX from substrates, e.g. soil or water.
XX
SQ Sequence 392 AA:

Query Match 6.9%; Score 152.5; DB 22; Length 392;
Best Local Similarity 21.6%; Pred. No. 1.6e-07;
Matches 89; Conservative 68; Mismatches 182; Indels 73; Gaps 16;

QY 9 LGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEAF--AFTLYGLAAIASAV 67
DB 22 MGMPMIMFY-----ALGILG-----PHLVADLGISROOLGMLTFSTGALAAISPMA 68
QY 68 SGVVAELITPQKTMIGFVLMCVFHLFLVFGGQANGGLILFVGIGLAIVPLFYSFI 127
DB 69 GALVORMG--RAGLI-----CFLLVGL--SFSIMAVLPGGGIVTALLLCGTA 114
QY 128 -----VVIHNVSESSSALGWYVAVSVGIGVAGSYIPSTPIPMGEMLTL--W 176
DB 115 QSLANPATNOALIAHSVPAKAGVGLKQSGVQASALAGVALPLVL--WNGRGALAAW 173
QY 177 LALAFCEPAGVIAIMISLRHVKTGGMHMLTPREKFAELSRVATLLYTRNIFLSSI--VR 234
DB 174 VPAALVMAALVITYWVPAKVSAP-----SLPLRVRGPRVWMLSIILMAIQ 216
QY 235 IINTSLFGEFAVIMPMFVDELGFTTSEMLOWAAFEFTTISNITGIVAEEKMGMYRI 294

```

Db      217 LCAGLALSEFMFLG-VYAAQIGVSVSTIGAMVSCFGAMGIIISRLVLLPTADLKROETLL 275
OY      295 RMEGCLGMAASSLAIFYMPOYFGHNWMMIPAIAGTGYAAVPMAAVFPALPEKHKA 354
Db      276 --LGVLFLIA-GLAALVMEANTQOHPMLGLGTGVTAASNAIAMSLLDGRFGA 332
OY      355 AISVYVLSAGMSNFLA---PATAVVL-----LPMESTIGVAVTALYLL 396
Db      333 AHSAGMLSTVGFGGFAVGPFAFGMFLAHSEGFPAAMLSLIGILVAGGLCLL 384

RESULT 13
ABBA8316
ID      ABBA8316 standard; Protein; 401 AA.
XX
AC      ABBA8316;
XX
DT      05-FEB-2002 (first entry)
XX
DE      Listeria monocytogenes protein #1020.
XX
KW      Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW      vitamin B12; bacterial infection; disease.
XX
OS      Listeria monocytogenes.
XX
PN      MO200177335-A2.
XX
PD      18-OCT-2001.
XX
PF      11-APR-2001; 2001WO-FR01118.
XX
PR      11-APR-2000; 2000FR-0004629.
XX
PA      (INSP) INST PASTEUR.
XX
PI      Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI      Dussurget O, Chetouni F, Nedjar H, Glaser P, Kunst F, Cossart P;
PI      Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI      Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI      Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L;
PI      Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI      Madueno E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J;
PI      Rose M, Voss H;
XX
DR      WPI; 2002-010914/01.
XX
PT      Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT      and prevention of Listeria and related bacterial infections, and
PT      related polypeptides
XX
PS      Claim 6; SEQ ID NO 1021; 192pp; French.
XX
XX      The present invention relates to the genome sequence of Listeria
CC      monocytogenes EGD-e (see AB03041). The genome sequence and fragments of
CC      it are useful for selecting probes and primers for detecting genes in L.
CC      monocytogenes and related organisms, and for studying genetic
CC      polymorphisms and other genomes. The present sequence is a protein
CC      encoded by the genome sequence of the present invention. Proteins
CC      expressed from the genome sequence are useful for raising specific
CC      antibodies, identification of L. monocytogenes and related organisms, and
CC      for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC      B12. The genome sequence and proteins encoded by it are also useful for
CC      selecting compounds that regulate gene expression and cell replication
CC      and modulate L. monocytogenes-related diseases. In addition, the genome
CC      sequence and proteins encoded by it are useful in pharmaceutical and
CC      vaccine compositions for the treatment or prevention of infections by L.
CC      monocytogenes and related organisms.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX

```

```

SQ      Sequence      401 AA;
Query Match      6.6%; Score 147; DB 23; Length 401;
Best Local Similarity 23.6%; Pred. No. 6.2e-07;
Matches 45; Conservative 40; Mismatches 88; Indels 18; Gaps 4;

OY      224 NRNIFLSIVRIINTSLGFAVIMPMFVDELGFTTSEMVLQWMAFFFTIPSNIFMG 283
Db      7 NKGFTLITLINFVLYVLLMYIIAVIAQDELNASLGEFGASGIYIIGTLARLYMK 66
OY      284 VAEKMGMRVIRMEGCLGMAASSLAIFYMPOYFGHNWMMIPAIAGT-----FVA 336
Db      67 KLELFGRKRYLR-FGILFELITTMAYLYMPT-LAIWEIIFRLNGFAFGFTSTATNAIVTA 124
OY      337 FVPMMAVFPALPEKHKAISVYNLSAGMSNFLAPALAVLLPMESTIGVAVTALYLL 396
Db      125 YIPNS-----RNGEGINYYGLSTSLAAIGPFIIGMLLSKTSFYTIITFIVILL 175
OY      397 AFVLCAFIRVE 407
Db      176 TALLCFYLPYK 186

RESULT 14
ABPA0749
ID      ABPA0749 standard; Protein; 393 AA.
XX
AC      ABPA0749;
XX
DT      24-JUL-2002 (first entry)
XX
DE      Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5594.
XX
KW      Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW      antibacterial; gene therapy.
XX
OS      Staphylococcus epidermidis.
XX
PN      US6380370-B1.
XX
PD      30-APR-2002.
XX
PF      13-AUG-1998; 98US-0134001.
XX
PR      14-AUG-1997; 97US-055779P.
XX
PR      08-NOV-1997; 97US-064964P.
XX
PA      (GENO-) GENOME THERAPEUTICS CORP.
XX
PI      Doucette-Stamm LA; Bush D;
XX
DR      WPI; 2002-381255/41.
XX
DR      N-PSDB; ABN93294.
XX
PT      Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT      polypeptide, useful for diagnosing and treating bacterial infections
XX
PS      Disclosure; SEQ ID 5594; 267pp; English.
XX
XX      ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC      frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC      given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC      antibacterial activity and can be used in gene therapy. The sequences
CC      can also be used in the diagnosis and treatment of bacterial infections,
CC      particularly S. epidermidis infections. The sequences can be used to
CC      screen for compounds able to interfere with the S. epidermidis life
CC      cycle or inhibit S. epidermidis infection.
CC      N.B. The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from the
XX      USPTO web site.
XX
SQ      Sequence      393 AA;

```



```

Db 282 IGVYLSNFQMGRTAMSADG---TAGEMLGSWTIIFYWMMWISWSPFVGMFLARISGRS 338
QY 294 IR-----WFGCLGMMASSLAIFYMPOYFGHNYW----- 321
Db 339 IREFIIGVLLVPAGVSTWPSIFG--GTAIVF-----EONGESIWDGGAEEQLFGLHAL 392
QY 322 ----MAMIPALGTFVAFVPMVAVPALPEPKH-----KGAATSVYNL 361
Db 393 PGQIMGITIMILGTFEFTTSADSASTVMGTMQHGQLEANKWYTAAGVATAAIGLTL 452
QY 362 SAGMSNFLAPAIIVLL---PW--FSTIGVIA 389
Db 453 LSGDNALSNLQWVTIYAATPFLFVYIGLMFA 484

```

Search completed: March 13, 2003, 16:53:13
 Job time : 36.7483 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:37:32 ; Search time 11.5828 Seconds
(without alignments)
1077.057 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211
Sequence: 1 MSRRNKQWLGIPHLIMGYI.....RVGPGFSSAPYTERALNIS 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1040.5	47.1	467	US-09-134-001C-3020	Sequence 3020, Ap
2	158.5	7.2	416	US-09-333-208-2	Sequence 2, Appl1
3	158.5	7.2	416	US-09-333-254-2	Sequence 2, Appl1
4	158.5	7.2	416	US-09-183-270-2	Sequence 2, Appl1
5	145	6.6	393	US-09-134-001C-5594	Sequence 5594, Ap
6	136	6.2	492	US-09-134-001C-4847	Sequence 4847, Ap
7	132.5	6.0	423	US-09-134-001C-5210	Sequence 5210, Ap
8	126.5	5.7	473	US-09-134-001C-3564	Sequence 3564, Ap
9	125	5.7	477	US-09-134-001C-3487	Sequence 3487, Ap
10	124	5.6	479	US-09-134-001C-4435	Sequence 4435, Ap
11	123.5	5.6	515	US-08-063-552-4	Sequence 4, Appl1
12	123.5	5.6	515	PCT-US93-05704-4	Sequence 4, Appl1
13	123	5.6	418	US-09-030-267-5	Sequence 4, Appl1
14	122	5.5	499	US-09-134-001C-5370	Sequence 5370, Ap
15	121.5	5.5	412	US-09-134-001C-3949	Sequence 3949, Ap
16	119.5	5.4	514	US-08-063-552-13	Sequence 13, Appl1
17	119.5	5.4	514	PCT-US93-05704-13	Sequence 13, Appl1
18	119.5	5.4	534	US-09-031-392-4	Sequence 13, Appl1
19	119.5	5.4	534	US-09-299-549-4	Sequence 4, Appl1
20	119.5	5.4	534	US-09-610-417-4	Sequence 4, Appl1
21	117.5	5.3	413	US-09-134-001C-5503	Sequence 4, Appl1
22	117	5.3	668	US-09-134-001C-3430	Sequence 5503, Ap
23	114	5.2	177	US-09-134-001C-4482	Sequence 3430, Ap
24	114	5.2	400	US-09-134-001C-2912	Sequence 4482, Ap
25	113.5	5.1	518	US-09-134-001C-4744	Sequence 2912, Ap
26	112	5.1	831	US-08-677-734A-11	Sequence 4744, Ap
27	112	5.1	831	US-09-097-053-11	Sequence 11, Appl1

28	110	5.0	429	US-08-677-049-5	Sequence 5, Appl1
29	110	5.0	516	US-08-356-340-4	Sequence 4, Appl1
30	110	5.0	516	US-08-786-555-4	Sequence 4, Appl1
31	108.5	4.9	521	US-08-063-552-2	Sequence 2, Appl1
32	108.5	4.9	521	PCT-US93-05704-2	Sequence 2, Appl1
33	108	4.9	462	US-08-898-976-2	Sequence 2, Appl1
34	108	4.9	462	US-08-898-976-4	Sequence 4, Appl1
35	107.5	4.9	662	US-09-134-001C-4074	Sequence 4074, Ap
36	107.5	4.9	1040	US-09-134-001C-5365	Sequence 5365, Ap
37	107	4.8	408	US-09-134-001C-4130	Sequence 4130, Ap
38	106	4.8	367	US-09-134-001C-5557	Sequence 5557, Ap
39	106	4.8	466	US-09-134-001C-3355	Sequence 3355, Ap
40	106	4.8	513	US-09-097-889-15	Sequence 15, Appl1
41	106	4.8	620	US-08-301-722A-2	Sequence 2, Appl1
42	105.5	4.8	381	US-09-134-001C-3743	Sequence 3743, Ap
43	105.5	4.8	398	US-09-134-001C-4353	Sequence 4353, Ap
44	104.5	4.7	493	US-09-031-392-10	Sequence 10, Appl1
45	104.5	4.7	493	US-09-299-549-10	Sequence 10, Appl1

ALIGNMENTS

```
RESULT 1
US-09-134-001C-3020
; Sequence 3020, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-1007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3020
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3020

Query Match      47.1%; Score 1040.5; DB 4; Length 467;
Best Local Similarity 48.6%; Pred. No. 6.5e-96;
Matches 195; Conservative 76; Mismatches 129; Indels 1; Gaps 1;

QY 10 GLPLHLIMGYITAAVEMTGDGPELAFSLSHYIKSLGFPAPASFAFTLYGLAALSAWVG 69
DB 55 GLPLSLMGFWATAVFTGTGIELAFSLRYVNDGSPDTATLLFYGLAALISSWVG 114
QY 70 VVAEITPQKTMIGFVLMCVFHVLPVFGIGQANVGLILFLFYGIRGLAYPLFYSFTIV 129
DB 115 VLAETGPRRIMLIGVAAVIVPHVLPFLFGLGQKTYLWVFGIRGLAYPLFYARMW 174
QY 130 IINWNSSENSALGAWVANYSVGIGVAGSYTFTPIPMGEKTMALAFCEAGVIA 189
DB 175 IAOVTPCARAASAMGWMSYCGIGLGNWIPSLISIRGFTINTLWVGFWAVALGMI 234
QY 190 MISLRHKHTTGHMHNLPKRFAPLNSAVTLTNRNIPFISITRIINTSLGFAVIM 249
DB 235 MYLVEKGACRPPDAVTLVLRKLTSLSGVITIERKRMFLIVLRITCNLSFGLPYILP 294
QY 250 MMEVD-ELGFTTSEMLOVMAAFETTTFSNIFWGIVAEGKGMWRVIRMFCLGMAASSLA 308
DB 295 LYLTVSEVGSMDQMLHLMGVMEFVVSIFTNVIMQIGDRIGMLIQHMFCICGALSLSLA 354
QY 309 FYTMDPTFGNINYMAMIPALITGFAAAYPMAAVPALEBKHAISVYNISAGMSNF 368
DB 355 FYVLDPGYGAHFETALTAIGFISVYAPVPMGAVFLALAPBEKGAAISAHNLAAGLSNP 414
```

```

QY      369 LAPRAIVLLPWFSITGVVATYATALLLAFVCAFIREEQ 409
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      415 MGPRIATLFIATLGIKGVWVIYAGLVAGAVLTFEPIRPQ 455

```

RESULT 2
US-09-333-208-2

```

Sequence 2, Application US/09333208A
Patent No. 6225089
GENERAL INFORMATION:
APPLICANT: Chen, Keyin K.
TITLE OF INVENTION: A Putative Solvent/Antibiotic Resistant Gene from
TITLE OF INVENTION: Pseudomonas Mendocina
FILE REFERENCE: CL-1160-C
CURRENT APPLICATION NUMBER: US/09/333,208A
CURRENT FILING DATE: 1999-06-15
EARLIER APPLICATION NUMBER: 09/163,270
EARLIER FILING DATE: 1998-10-30
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 416
TYPE: PRF
ORGANISM: Pseudomonas mendocina KR-1
US-09-333-208-2

```

Query Match	7.28;	Score 158.5;	DB 4	Length 416;
Best Local Similarity	22.78;	Pred. No. 9.8e-08;		
Matches 88;	Conservative 58;	Mismatches 164;	Indels 77;	Gaps 21

Qy	20	IAIVFMFG-----DGEFLAFLSHYIKSLGTPPAEASFATLYG-----LAAAL	63
Db	66	VAVVVLTLGLVAANYGFGIYLFADLPDM--QASLSEFGFSYVGTITASQNLGFLLCAML	1222
Qy	64	SANWSGVAAEITIQ---KTYMLGFEVLMCFVHYLFVFGIGQANYGLILFLFTIGRLAYP	1200
Db	123	AVW-----LTPKVGSGRMIAASGAVCALALLILPLSNTEVIGLLTL--LACTAAT	1772
Qy	121	LELYSEIVVLIHANYRENSSSALGAMVANYVSGIGVAGSYSPSTPIIMEMGTMLALA	1800
Db	173	VFV-PMDVIARVAYAIRGRGLAMGLVSSGTSYGAINSLPLPYA--PQGBMRSVMWVVG	2229
Qy	181	-FCEAGGVIAAISLRHVKTPGHMNLTPREKFAELSAVTLTYNRMILFSLIYR-----	2340
Db	230	LLTLAMTYLVLYVL--KRGLLDQALPASAATATIDDEAS-----SSGLSGLRIPVYLL	2800
Qy	235	IINTLSLGEFAVIMPMFVD-----ELGFTTSEMLOQVMAAFETTTIFSNIWFQVIAEKMG	2899
Db	281	IMSNNFLIGFATPFQFUTLSYLRTELGEFDQVYQAQVAVIGFVGMFAGLAVGMLSDRIG	3400
Qy	290	WMRYIRW--FECCLGMAASSLAFYTMPOYFGHNTYMMAMIPALIGTVPAAVPMANYPAL	3473
Db	341	-LRFAMTLVYCCVYTA--LIEVVQPS-GH-W-----PLVAALFSTAFYPIFGLIPAY	3899
Qy	348	EPK---HKGAISVYNL-----SAGM	365
Db	390	VSKLASSAMAVSIFGLIANVMOGGSGM	416

```

: RESULT 3
: US-09-333-254-2
: Sequence 2, Application US/09333254A
: Patent No. 6235882
: GENERAL INFORMATION:
: APPLICANT: Chen, Kevin K.
: TITLE OF INVENTION: A Gene Encoding a Putative Efflux Protein for Solvents/
: TITLE OF INVENTION: Antibiotics in Pseudomonas Mendocina
: FILE REFERENCE: CL-1160-D
: CURRENT APPLICATION NUMBER: US/09/333,254A
: CURRENT FILING DATE: 1999-06-15
: EARLIER APPLICATION NUMBER: 08/961,738
: EARLIER FILING DATE: 1997-10-31
: NUMBER OF SEQ ID NOS: 6

```

```

: SOFTWARE: Microsoft Office 97
: SEQ ID NO 2
: LENGTH: 416
: TYPE: PRT
: ORGANISM: Pseudomonas mendocina KR-1
US-09-333-254-2

```

Query Match	7.2%;	Score 158.5;	DB 4	Length 416;
Best Local Similarity	22.7%;	Pred. No. 9.8e-08		
Matches 88;	Conservative 58;	Mismatches 164;	Indels 77;	Gaps 21

QY	20	IAIAVFMTQ-----DGEFLAFHSYKSLGFTPAEASFAPTLVG-----LAAAL	63
Db	66	VAVVVLTLGLVAATYGGGCIYLPQLVPDM---QASLGFGSYGVTITASAQLGFLLCAML	1222
QY	64	SAMVSGVAEITTPQ---KTMILGEVLACVPHNLEFLVGGJGQANYGLILLPYIGRLATP	1200
Db	123	AWV-----LFPKVGSGRMIAASGAVCALALLILPLSNTFEVIGLLTL--LAGTAAT	1722
QY	121	LELYSEFLVYIIHNVRSESSALGMYANVSVGIGVAGSYIPSTPIIMEXGTLMLALA	1800
Db	173	VFV--PMVDVIARVAVAYRGLAMGLVSSGTSYGVAINSLVPIYA--POEEMRSVMWVG	2299
QY	181	FCGAGGVIMISLRHNTKPGHMNNLPRKRPALVRAVTLVTNRNIEFLSYR-----	2344
Db	230	LTLMTAVLVLVVL---KRGGLGQALSPAPTLTDEAS-----SSGLSGLRPVLL	2800
QY	235	IINTLSFGEFVAVIPMMFVD-----ELGFTTSEVLQWMAAFETTLTIPSNIFWGIVAEKM	2899
Db	281	IMSNNFLIGTATPFQYIYLSYLRTELGFDVQYTAQWMAVIGVGMNAGLAVGMLSDRTG	3400
QY	290	WMRVYRN--FCGLGMASSLAFFYMPQYFGHNWMMAMIPALIGTVAALFPMMAVFPAL	3477
Db	341	LRFAMTLVYGCVTAA--LTFVQPS--GH--W-----PLVAVLFTSTAYPIFGILPAY	3899
QY	348	EPK---HKGAISVYNL-----SAGM	365
Db	390	VSKLASSSMAVYSIFGIANWOGGGM	416

```

RESULT 4
US-09-183-270-2
Sequence 2, Application US/09183270B
Patent No. 6410265
GENERAL INFORMATION:
APPLICANT: Chen, Kevin K.
TITLE OF INVENTION: A Putative Solvent/Antibiotic Resistant Gene
FILE REFERENCE: CL-1160-A
CURRENT APPLICATION NUMBER: US/09/183.270B
CURRENT FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: 08/961,738
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 416
TYPE: PRS
ORGANISM: Pseudomonas mendocina KR-1
US-09-183-270-2

```

Query Match	7.2%	Score 158.5;	DB 4;	Length 416;
Best Local Similarity	22.7%	Pred. No.	9.8e-08;	
Matches	88;	Conservative	58;	Mismatches 164; Indels 77; Gaps 21.
OY	20 IAINVEMTG-----DGEFLAFLSHIKSLGTFPAASAFITLYG-----LAAAL	63		
	: : : : : :			
Db	66 VAVVVLLTGLVAATYGFGIYLFAQLVPDM---QASLGGFSVSXGITASAOLGFLLCAML	122		
OY	64 SAWSGVAALIEIPO--KTMLIGFELWCVFHVLTFLVGLGANGLILFFYGIKGLAMP	120		
	: : : : : :			
Db	123 AAW-----LTPRVGGGRMIASGAVALCALALLPLSSNTFYGVILLTL--LNGTAFT	172		

; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5210
 ; LENGTH: 423
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-5210

Query Match 6.0%; Score 132.5; DB 4; Length 423;
 Best Local Similarity 22.7%; Pred. No. 4e-05;
 Matches 90; Conservative 60; Mismatches 123; Indels 124; Gaps 23;

QY 90 VFHVLFLVFG-----LGOAN-----YGLILFLYGRGLAFLPLYS 125
 Db 67 VFPCVFLIFSPIGLKGRKNDKPEFTISFMALFSAGMIGLVFY---GAAPMAHFA 123
 QY 126 -----FTVLIHNVRENSSSALGW---YMAVSVGIGVAGSY-----IPST 165
 Db 124 APPTADEPTTKAYTESLR-----STFFHMGFHAMAIYGV-VALLALAYSQFRKGEPLISRT 178
 QY 166 I-PIHGE-----KOTMLALA-PCFAGCVIAMISLRHVKTRGHMNL--TPREKFAE-L 214
 Db 179 LRPLLGKVEGPIGTLDIVLSEFATLVGVAVSLGMALQINGGLHLYFGVPMNTFVGGII 238
 QY 215 SRAVTLTY-----TNRNIFLSIYRIINTLSLFGFAVIMPMFVDELGFT 259
 Db 239 IYVVTILFISANSGLSKGYOYLSNLMIGCTILMIY-TLIVGPTVILNNM----- 289
 QY 260 TSEMLOVMAAFETTTISNIFMGIVAEKMGMRVIRWEGCLGAASSLAFYMPQYFGHN 319
 Db 290 TSTGSLNSFLFNSFDTALNG---OKRDM-----STWTLTYW-----G 327
 QY 320 YMAAMITAIKGTVA-----AFVPMAPPALEPKHKAISVYNLSAGM--SNF 368
 Db 328 WMLSWSEFV--GVFIAYSKGRSIREFISGLVLPALVSEFVSGVIGIEAGKDSGL 385
 QY 369 LAPAIIVLLPWFSTI--GVFIATYALYLAFVLCAF 403
 Db 386 FKMSPEQGLGVFNHPIGLIVSLIALLIASLCLYF 422

RESULT 8
 US-09-134-001C-3564
 ; Sequence 3564, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3564
 ; LENGTH: 473
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3564

Query Match 5.7%; Score 126.5; DB 4; Length 473;
 Best Local Similarity 20.3%; Pred. No. 0.00019;
 Matches 100; Conservative 73; Mismatches 169; Indels 151; Gaps 26;
 QY 1 MSRNKQWLG-LPLHLIMGYIAIVPMGDDGFELA--FLSHYKSKGFTPAEASF---A 53
 Db 30 LSVQRKLMKNFQAFVVFVYVYAMYLIRNNEKAAOPLK---EETGLTTLLEIGYIGLA 86

QY 54 FTL-YGLA-AALSAWVG-----VVAEITPQKTMIGFVLCVHVLEVLVEGLG 101
 Db 87 FSIITYGLKGTILGYFVQGRNTRKRIISFLILSAT---VIMGFVL-----SYFG-- 133
 QY 102 QANTGLILFYGIRGLAFLPLYSFIYVLIHNVSENSSA-----LGYWMA 148
 Db 134 -SVMGILLIVLMNG-----IFQSVGPASYSSTISRMARTRKRGYLGFWNT 179
 QY 149 VYSGIVAGSYRPSFTPIPMGEWGLMLALAFCFAGCVIAMISLRHV----- 196
 Db 180 SHNIGAIAGV-----ALMGANTF-FHGNVGMFTFSPVIALIIGIVLFI 225
 QY 197 -----KTRGHMHN-----LTPREKFAE--LSRAVTLTYNRIPLSSI 232
 Db 226 GKDDPEELGNWRABEIEEPIQENIDISOGMTKDKFKYILGNPVIWILCISNVF-YI 284
 QY 233 VRIINTLSLGFVAVIMPMFVDELGFTTSEMLOVMAAFETTTISNIFMGIVAEKMGMR 292
 Db 285 VRI-----GIDMMAELVYSEHLHFNKGDVNTIFFEIGALVASLGMYSIDLKGR 337
 QY 293 VIRMFGCLGAASSLAFYMPQYFGHNYMAMIPALIGFVAFVMAAVFPALE--P 349
 Db 338 AIVAGCMFTTEFVLETTN---ATSVTKVNISLFLGALI--FGQLIGVSLTGFPV 391
 QY 350 KHKGAISVYNLSAGMSNF-----APAIIVLLPWFSTIGV-----VIAYT 391
 Db 392 KN--AIVSANGMGSPFAYLFGDSMAKVLAALIDPFRNGIINTIRGTYLGSMTDVFIFYV 448
 QY 392 ALYLAFLVLCAFI 404
 Db 449 ALF-LGMILLAIY 460

RESULT 9
 US-09-134-001C-3487
 ; Sequence 3487, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3487
 ; LENGTH: 477
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3487

Query Match 5.7%; Score 125; DB 4; Length 477;
 Best Local Similarity 22.2%; Pred. No. 0.00027;
 Matches 107; Conservative 74; Mismatches 172; Indels 128; Gaps 26;
 QY 1 MSRNKQWLGILPLHLING-----YI-AIVAFMGD-----GEELA 34
 Db 50 IETTAQWMLITGFMVINGIMIPLTAFLMDKYSTRNHLYIFGMAIFLGSIIVAASPTTIL 109
 QY 35 FLSHYIKSLG---FTPAESFAFTLY-----GLAALSAWVGVAIIRPQKTMILGEV 86
 Db 110 MISRIIDAIAGILPLPMOFTVFTLPAEORGFAMGL---AGVYV-----QSPATIGPT 160
 QY 87 LMCVFHVLFLVFGIQAQNYGILILFYGIRGLAFLPLYSFIYVLIHNVSENSSS---A 142
 Db 161 LTGLFVDLF-----SMRMPEFLVSAIAVAFLTGFFV-----ENNTKTKDIV 203
 QY 143 LGWYAVYSV-GIGV---AGSYIPSTPIPMGEWGLMLALAFCFAGCVIAMISLRHVKT 198

```

Db      204 LDKISVYSTFGGLILFAFSSVSFGITSLPIYVTEVGLIA-----IIIIETQQLK 257
Oy      199 PGHMHLTREKFAELSRVAVTLTYTRNIFLSIVRIINTLSLFGFAVIMPMFVELGE 258
Db      258 KHPLLMR-----VEKNKVPFLSVSSMLVYITWVSPALLIPYI--QTGI 301
Oy      259 TTESEMLQVAAFEFTTII--FSNIFMGIVAEKMG--NMRYIRMEFCGLMAASSLAFYTMPQ 315
Db      302 GGSALLSGVVLPGAVINGLTWYTKIRDKHGIKVLIPGF-----ILLISMTFLXSELT 357
Oy      316 FGHNYWMA-----IPALAGTFVAAFEVMAAV--FPALER--KH-----KG 353
Db      358 TGTPEYFVILYVYTRIMIALGLLV---MPLNTVGLNALESDDYSHGTAIMNSLRILAGMG 414
Oy      354 AAISSVNSAGSMNLA--PATAVALLPWFSTI--GVVIAY---TALYLAFCATIRRE 407
Db      415 TVAVSYTILSVAKQYVASHSTMSKMLTQEAIVHGIDVAFITTVLIIIGLILAFIKKE 474
Oy      408 Q 408
Db      475 K 475

```

```

RESULT 10
US-09-134-001C-4435
; Sequence 4435, Application us/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lyna Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4435
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4435

```

```

Query Match      5.6%; Score 124; DB 4; Length 479;
Best Local Similarity 19.5%; Pred. No. 0.00034;
Matches 86; Conservative 78; Mismatches 190; Indels 88; Gaps 20;

```

```

Oy      13 LHLIMGYIAIVMTGDFELAFLSHYIKSLGFTTPAEASFAFTLGLAALASWVGVA 72
Db      59 LGIFITG--AGYLLKRNKSLAMP--LIEQGIKGLGIALSAVSIAVGFSGFYMGIVS 114
Oy      73 EITTPKTMILGFVLMCVFHV--FLVFGIGANYGLILFY--GIRGLAYP---LFLY 124
Db      115 DRSNAMFLPLGLVLAITINLLGIFPFSSITIMEFLVGMQGMWMPSGRVLVH 174
Oy      125 SPLVYIIHVRSESSSALGWTAVYVSGIGVAGSYIIPSTPIIMGEMGLMLALAFCA 184
Db      175 GFSV-----SERGSKTSIMNVA--HNVGQ-----LMPIATWGISMTALYN 214
Oy      185 GG-----VIAMISLRHVKTPEGHMHNLPREK-----AELSRATLLY 222
Db      215 FGLKFGVGYIYPALLAIIIAIFSYILIRDPQSGLPIDYQKDYATSTKOTIETEL 274
Oy      223 TNRNIFLSIVRIINTLSLFGFAVIMPMFVELGFTTSEMLOY-----NA 268
Db      275 TTREILFKYVLN---NMKWWAIAFTNIFVYVRGVLDWMAPTYLSEKHPLDSAGMA 329
Oy      269 APFF--TTTFSNIFMGIVAEKM--GMRYIRMGCGICMAASSLAFYTMPQFSGHNYWMA 324
Db      330 YFLYEWAGIPGTLGCLSDKLFKGRGRGAPGFFFMGLGVITIFILYMLNPP--GH--AWLDN 386

```

```

Oy      325 IPALAGTFVAAFEVMAAVPEALPEPKHGAISVUNYNSAGSMNLAIPATAVALLPWFSTI 384
Db      387 LSLIGIGFLILYGV--MLIGIQLDLYVPKKAAGTAAGLT--GLFGYLGAVYMANIV-----L 439
Oy      385 GVVIAYTALYLAFCATIRV 406
Db      440 GFVQHFQGMH--IGEVLLTVISI 460

```

```

RESULT 11
US-08-063-552-4
; Sequence 4, Application us/08063552
; Patent No. 5688936
; GENERAL INFORMATION:
; APPLICANT: Edwards, Robert H
; TITLE OF INVENTION: Vesicle Membrane Transport Proteins
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/063,552
; FILING DATE: 19930514
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Farber, Michael B
; REGISTRATION NUMBER: 32,612
; REFERENCE/DOCKET NUMBER: 9067-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-063-552-4

```

```

Query Match      5.6%; Score 123.5; DB 1; Length 515;
Best Local Similarity 21.8%; Pred. No. 0.00042;
Matches 89; Conservative 63; Mismatches 157; Indels 99; Gaps 23;

```

```

Oy      58 GLAALASAWY-----GVVAETIPPKTMILGFVLMCVFHVLFVFGIGANYGLIILL 110
Db      133 GLLFASKATVQLTNPFGLITRNIGYIPMPAGF---CIMFTSVFAF--SSSYFLLI 188
Oy      111 FYGIRGLAYPLFLXSFVIYIIHVRSESSSALGWTAVYV-----VGIGVAGSYI- 181
Db      189 ARSLQIG-----SSCSYVAGMGMLASVYTTDDEBKRGPGIALGGLAMG 232
Oy      162 ----PSF--TPIIMGEMGLMLALAFCFAGVYIAMISLRHVKTPEGHMHNLTPREKFAEL 214
Db      233 VLVGPPGSLYLFVGVKTAFLVLAALVLDGALQLFVLQPSRVQESQKGP----- 285
Oy      215 SRAVTLTYTRNIFLSIVRIINTLSLFGFAVIMPMFVELGFTTSEMLOYAAFEFTT 274
Db      286 --LTTLLKDPYILIAAGSICFAN---MGIAMLEPALPIMMETMCSRKWQLGVALPAS 339
Oy      275 IF---SNIFMGIVAEKGMRYIRNF--GCLGMAASSLAFYTMPQFSGHNYWMAIP--- 326
Db      340 ISYLTGNTF--GILAHKMG-----RWLCALLGVYIVGISILCIP--FAKNYGLIAPNFG 391

```

QY 327 -AALGTFVAAPMAAFVPALEPKHGAISY---NLSAGSNFLAP---AIAVVL- 377
DB 392 VGFAIGAVDSSMPIMGYL--VDLRHVSYGVYAIADVAFCMGYAIGPSAGAIKAIG 449
QY 378 LPWFST-IGVV-IAYTALYLALFVLCAFIRVEOPGSSAPVTEKALNI 423
DB 450 FPMMLTIIGIIDIADFAP-----LCFFLR-----SPPAKEKKMI 483

RESULT 12

PCT-US93-05704-4
; Sequence 4, Application PC/TUS9305704
; GENERAL INFORMATION:
; APPLICANT: Edwards, Robert H
; TITLE OF INVENTION: Vesicle Membrane Transport Proteins
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05704
; FILING DATE: 19930611
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Faiber, Michael B
; REGISTRATION NUMBER: 32,612
; REFERENCE/DOCKET NUMBER: 9067-1PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-05704-4

Query Match 5.6%; Score 123.5; DB 5; Length 515;
Best Local Similarity 21.8%; Pred. No. 0.00042;
Matches 89; Conservative 63; Mismatches 157; Indels 99; Gaps 23;

QY 58 GLAAALSAMWS-----GVVAELITTPQKTMIGFVLMCVFHVLFVFGSGQANVGLIL 110
DB 133 GLIFASAKATQQLTNPIGILTRIGYPIPMAGF--CLMISTVWFAF-SSSYAFLLI 188
QY 111 FYIRGLAIPLFLXSLFVLIHNVRSSENSSALGMWAVS-----VGIGVAGSYT- 161
DB 189 ARSLGGG-----SSCSVAAGMGMLASVYTTDEERGKPMGIALGGLAMG 232
QY 162 ----PSF---TIPMGEMGTMLALAFCCFAGVYAMISLRHVKTPTGMMHULTPREKFAEL 214
DB 233 VLVGPPGSLVLEFVGKTAFLVLALVLLDGAIOLEFVLPKSRKQPSQKTP----- 285
QY 215 SRAVTLTYNRNIFLSIVRIINTLSLGFVAVIMPMFVDELGFTTSEMLQVMAAFETT 274
DB 286 --LTTLTKDPIYLIIAAGSICFAN---MGIAMLEPALPIMMERMCKRQKQDGVAFILPAS 339
QY 275 IF-----SNIRGVIAEKMGMRIYIRNF-GCLGMAASSLAFYVPOYFGHNYWMAAMP--- 326
DB 340 ISYLVIGINIF-GILAHKMG-----RWLCALDGMVIVGISILCIP--FAKNIYGLIADNFG 391
QY 327 -AALGTFVAAPMAAFVPALEPKHGAISY---NLSAGSNFLAP---AIAVVL- 377

DB 392 VGFAIGAVDSSMPIMGYL--VDLRHVSYGVYAIADVAFCMGYAIGPSAGAIKAIG 449
QY 378 LPWFST-IGVV-IAYTALYLALFVLCAFIRVEOPGSSAPVTEKALNI 423
DB 450 FPMMLTIIGIIDIADFAP-----LCFFLR-----SPPAKEKKMI 483

RESULT 13

US-09-030-267-5
; Sequence 5, Application US/09030267
; Patent No. 6162632
; GENERAL INFORMATION:
; APPLICANT: Maloney, Peter C.
; APPLICANT: Keietsu, Abe
; APPLICANT: Zhong-Shi, Ruan
; TITLE OF INVENTION: OxIT SEQUENCE AND ITS USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave.
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,267
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/038,345
; FILING DATE: 27-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5789-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 31751
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-030-267-5

Query Match 5.6%; Score 123; DB 4; Length 418;
Best Local Similarity 20.6%; Pred. No. 0.00035;
Matches 98; Conservative 67; Mismatches 139; Indels 172; Gaps 23;

QY 17 WGYIAIAVEM---TGDFELAPLSHYIK-SLGFTPAEASFATLYGLAALASAMWSGV 71
DB 17 WFLYLVLLVLMCMISGVQVSMITYANPKDNLGAVLAOVTAFTLSQVIOAGSQPGGYF 76
QY 72 AEITTPQKTMIGFVL---WCVFHV-----LFLVFGSGQANVGLILTYGT- 114
DB 77 VDKFGPRIPIMFGAMVLAQWTEFMGWDVSPALYALTTLGAGVGY---YGIAMNTANR 133
QY 115 -----RGLA-----YPLFYSFTVLIHNVRSSENSSALGMWAVSYGVIGVAGSYIP 162
DB 134 WFPDQKRLAGGFTPAAGGLGLVLPPLPLISSVLAYE-----GVGAAPFYTG 178
QY 163 SFTIPMGEMGTMLALAFCCFAGVYAMISLRHVKTPTGMMHULTPREKFAELBRAVTLTY 222
DB 179 LI-----MGIILITLIAFV-----IRFPG-----QOGAKKOIVV 206
QY 223 TNRNIFLSIVRIINTLSLGFVAVIMPMFVDELGFTTSEMLQVMAAF----- 271

Db 207 TDKDNSEGMRL-----TFQFWLMTAFESVNEGGLLVA 241
 QY 272 -----FTTFSNT-----FMGIARMKG--MMRIYRMGCGLMAA 304
 Db 242 NSVYGRSGLAGVLTIGVSIONLFGCGRPFGVSDIGYKTMASV-----FGINA 296
 QY 305 SSIAFYMPQYGHNYMMAMPAIALGTEVAAVFPMAAFPALEPKHKAISVYN-----360
 Db 297 VVALFPTIALGDVAFIML-AIAFTWGSY-----ALPSTNSIDIFGTASARNVGF 351
 QY 361 --LSAGMSNF--LAPAIIVLLPSTIGVYIATYATLALLAVLCAPT--RVEQP 409
 Db 352 WAKATASIFGGGLGAIAIATNF--GWNATFLLTAITS--FIAPALATFVLRMRGR 403

RESULT 14

US-09-134-001C-5370
 ; Sequence 5370, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5370
 ; LENGTH: 499
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-5370

Query Match 5.5%; Score 122; DB 4; Length 499;
 Best local similarity 21.3%; Pred. No. 0.00057;

Matches 92; Conservative 60; Mismatches 169; Indels 110; Gaps 19;

QY 17 WGIYIAIVEMTGDGFELAFSLHYIKSIGFTTPARASFAFTLLYGLAALSAWVGVAEITP 76
 Db 36 WG---AFILPGD-----WIKSG--PIASSIGIVIGALLMILIAVSAGALVERKP 80
 QY 77 PQKTMILGFVLMGVPHVLIVFELGOANYGLILFLYG-----IRGLATVLPFLXPIVVI 130
 Db 81 -----VSGAFAFSFLSFGRYVSFFSSWFLTFGYVCVALMATAFSLIKFLPNV 131
 QY 131 IHNVRENSSSALGWYAWVYVSGIGVGYIPSE--TIPINGEGTLMALALAFCEAGVIA 189
 Db 132 LNWGKLYTAA-----WDVYITELIATVTLIYFMILITIGASVSGSLQYFFCA--MVL 184
 QY 190 MISLRHVKTPGHHM-NLTPREKFAELSRATVLTNTNRIPLASSIVRIINTLSLFGFVIM 248
 Db 185 VALMFIQSFSSHSFSLHLEPLASVDK-----WFGSLIMVS-----IA 225
 QY 249 PMMFV--DELGFTTSEWLOMAAFFTTIPSNIFFMGIVAAKGMGMRIYRFGCLGMASS 306
 Db 226 PMAYVGFNDINIPOTAEFE--NESPKNTEKLIVSLLAASLTYYVVALTYTGMISTQATS 280
 QY 307 LAFYIMPQYGHNYMM-----AMIPATALGTPVA--AFVP-----339
 Db 281 L-----NGNLMLTGAVTODAFGTIGLAVLAVAILIMIGITGLNGFLMSSSKRLFSMG 331
 QY 340 -----MAVFPALPEKHKAISVYNLSAGMSNFLAPAIIVLLPW--FSTIGVVIATYA 392
 Db 332 RSGIMPTVFESKLSKHTPYVAIIFLVA--VSLIAPLGRALTWIVDMSTGVSIAYFI 389
 QY 393 LYLAFVLCAP 403
 Db 390 TCISSATKLSF 400

RESULT 15

US-09-134-001C-3949
 ; Sequence 3949, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3949
 ; LENGTH: 412
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3949

Query Match 5.5%; Score 121.5; DB 4; Length 412;
 Best local similarity 19.0%; Pred. No. 0.00049;
 Matches 74; Conservative 81; Mismatches 168; Indels 67; Gaps 17;

QY 39 YIKS-LGFTP-----AESFAFTLYGLAALSAWVGVAEITIPQKTMILGFVLMCV 90
 Db 47 YILSYNLEFTVYSIAIATRSFAFSIHFIADATNFVIGFLKFSKLVISGFL--A 104
 QY 91 FHVLFVFGIGQANYGLILFLYGR-GLAY-PLFLYSFIVYIHNVSSENSSALGWYMA 148
 Db 105 FIFLFLVIMFPASP--IIFSAIMGLIAYSPIM-----VIMLSVDERNRKRGOMGVYR 157
 QY 149 VYVSGVAGSYIPSEFTIPMGEMGLMLALAFCEAGVYIAMSLSRHVKTGPHMHLTPR 208
 Db 158 SWLGL-LVGMVIMNMLIKHPTRFALMALVVLIAWLVYFVNI-----MLTYNTK 209
 QY 209 EKRAELSRATVLTNTNRIPL-----SSIVRIINTLSLFGFVIMPMFVDELGFTP 260
 Db 210 PVKQOLKQIVD--TORHLILPFGILLOGAIALVALPILKVIATQYVK---VSTVGTYV 263
 QY 261 SEMIQ-----VMAAFETTIPSNIFFMGIVAAKGMGMRIYRFGCLGMASSIAFYMPQYF 316
 Db 264 AIIIGIGCAFSLFLSKIIDN-----NSKGFMYGVIISGFI-----LYTILIF 307
 QY 317 G---HNYYMMAMIPALGTFVAAPMAAVFPA--LEPKHKAISVYNLSAGMSNFLA 370
 Db 308 GLSTITNITYIYMAIGLFIGLMYGILLPAMNTFMAGHINPNEDEETGCVFNSVGGFGSMIG 367
 QY 371 PAIAVILLPWFSTIGVYIATYATLALLAVL 400
 Db 368 PLVGGILTQFTNMLNNTFESAMIFLALAV 397

Search completed: March 13, 2003, 16:42:19
 Job time: 13.5828 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:41:12 ; Search time 11.9338 Seconds
(without alignments)
1637.624 Million cell updates/sec

Title: US-09-802-208B-5

Perfect score: 2211

Sequence: 1 MSRRNKQWMLGPLHLIMGYI.....RVGEQGFSSAPYTEKALNIS 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications-AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*

7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2211	100.0	444	9	US-09-802-208B-5
2	580.5	26.3	425	9	US-09-738-626-3627
3	143.5	6.5	595	9	US-09-738-626-6943
4	142	6.4	463	9	US-10-122-466A-11
5	139	6.3	417	10	US-09-815-242-10592
6	138.5	6.2	449	9	US-09-738-626-6555
7	136.5	6.0	378	10	US-09-815-242-10467
8	132.5	5.9	701	9	US-09-738-626-6796
9	130	5.9	453	9	US-09-738-626-6877
10	129.5	5.8	450	9	US-09-738-626-4815
11	128	5.7	429	10	US-09-922-501-10
12	126	5.7	407	10	US-09-815-242-11180
13	125.5	5.6	448	9	US-09-738-626-6795
14	124.5	5.6	472	9	US-10-024-623-29
15	123.5	5.6	663	10	US-09-815-242-14080
16	123	5.6	663	10	US-09-815-242-10070
17	122	5.5	462	9	US-09-738-626-5120
18	121.5	5.5	467	9	US-09-738-626-3974
19	121.5	5.5	467	9	US-09-738-626-3974

20	121	5.5	450	10	US-09-741-669-385	Sequence 385, App
21	120.5	5.5	427	10	US-09-922-501-2	Sequence 2, Appl1
22	120	5.4	426	10	US-09-922-501-4	Sequence 4, Appl1
23	120	5.4	551	9	US-09-738-626-4431	Sequence 4431, Ap
24	119.5	5.4	454	10	US-09-729-674-140	Sequence 140, App
25	119.5	5.4	469	10	US-09-925-302-627	Sequence 627, App
26	119.5	5.4	534	9	US-09-981-947A-4	Sequence 4, Appl1
27	119.5	5.4	649	10	US-09-815-242-5333	Sequence 5333, Ap
28	119.5	5.4	662	10	US-09-815-242-12343	Sequence 12343, A
29	119	5.4	526	9	US-09-738-626-5712	Sequence 5712, Ap
30	118.5	5.4	451	10	US-09-815-242-5079	Sequence 5079, Ap
31	118	5.3	513	9	US-09-738-626-6860	Sequence 6860, Ap
32	116	5.2	476	9	US-09-946-763-2	Sequence 5741, Ap
33	115.5	5.2	658	10	US-09-815-242-11824	Sequence 11824, A
34	114.5	5.2	395	10	US-09-815-242-13892	Sequence 13892, A
35	114	5.2	497	9	US-09-738-626-6734	Sequence 6734, Ap
36	113.5	5.1	464	9	US-10-024-623-28	Sequence 28, Appl
37	113.5	5.1	549	10	US-09-946-763-2	Sequence 3926, Ap
38	113.5	5.1	549	10	US-09-946-763-2	Sequence 2, Appl1
39	113	5.1	372	9	US-09-860-670-106	Sequence 106, App
40	112.5	5.1	391	10	US-09-815-242-11912	Sequence 11912, A
41	112.5	5.1	468	10	US-09-815-242-5144	Sequence 5144, Ap
42	112.5	5.1	487	9	US-09-738-626-4875	Sequence 4875, Ap
43	112.5	5.1	501	9	US-10-051-909-38	Sequence 38, Appl
44	112	5.1	403	9	US-09-738-626-6610	Sequence 6610, Ap
45	112	5.1	417	10	US-09-815-242-10165	Sequence 10165, A

ALIGNMENTS

RESULT 1									
US-09-802-208B-5									
1	Sequence 5, Appl1	US/09802208B							
2	Publication No.	US20030041352A1							
GENERAL INFORMATION:									
3	APPLICANT:	Parrot, Wayne							
4	APPLICANT:	LaFayette, Peter							
5	APPLICANT:	Kane, Patrick							
6	TITLE OF INVENTION:	Arabidol or Ribitol As Positive Selectable Markers							
7	FILE REFERENCE:	UGA-855R							
8	CURRENT APPLICATION NUMBER:	US/09/802, 208B							
9	CURRENT FILING DATE:	2001-03-08							
10	NUMBER OF SEQ ID NOS:	5							
SOFTWARE: PatentIn version 3.0									
11	SEQ ID NO:	5							
12	LENGTH:	424							
13	TYPE:	PRT							
14	ORGANISM:	Escherichia coli							
US-09-802-208B-5									
Query Match	100.0%	Score 2211:	DB 9:	Length 424:					
Best Local Similarity	100.0%	Pred. No. 2.5e-181:							
Matches 424:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:	
1	MSRRNKQWMLGPLHLIMGYIAIVFMTGDFELAFSLHYKSGFPFAPASPAFTLVGLA	60							
1	MSRRNKQWMLGPLHLIMGYIAIVFMTGDFELAFSLHYKSGFPFAPASPAFTLVGLA	60							
61	AALSAMVSGVAAEITIPQKTMILGFLVLCVFHVLFVFGGLGOANYGLILFYGIRGLAYP	120							
61	AALSAMVSGVAAEITIPQKTMILGFLVLCVFHVLFVFGGLGOANYGLILFYGIRGLAYP	120							
61	AALSAMVSGVAAEITIPQKTMILGFLVLCVFHVLFVFGGLGOANYGLILFYGIRGLAYP	120							
121	LFPLSYFVLIHNVRSNSSLALGMYAVYVSGVAGVSTPSTTIKMGMLTLALA	180							
121	LFPLSYFVLIHNVRSNSSLALGMYAVYVSGVAGVSTPSTTIKMGMLTLALA	180							
121	LFPLSYFVLIHNVRSNSSLALGMYAVYVSGVAGVSTPSTTIKMGMLTLALA	180							
181	FCFAGVYIAISLRHVTPGHHMLFPRKFAELSRVTLTLTYTRNFSLSYAIIWTL	240							
181	FCFAGVYIAISLRHVTPGHHMLFPRKFAELSRVTLTLTYTRNFSLSYAIIWTL	240							
181	FCFAGVYIAISLRHVTPGHHMLFPRKFAELSRVTLTLTYTRNFSLSYAIIWTL	240							
241	LFGRFVIMPMKVFDELGFTTSEMLOVAAFEFTTIFSNIFWGIVAEKMGMRVIRMGCL	300							
241	LFGRFVIMPMKVFDELGFTTSEMLOVAAFEFTTIFSNIFWGIVAEKMGMRVIRMGCL	300							
241	LFGRFVIMPMKVFDELGFTTSEMLOVAAFEFTTIFSNIFWGIVAEKMGMRVIRMGCL	300							

```
Db 241 LFGAVIMPMEFVDELGTTSENIQVMAAFETTTIFSNIFGVIYAEKKGMKRVIRMECL 300
Qy 301 GMAASSLAFFYMPQYFGHNWMMAMIPALIGCTFVAAPVMAAVPALPEPKHGAISYVN 360
Db 301 GMAASSLAFFYMPQYFGHNWMMAMIPALIGCTFVAAPVMAAVPALPEPKHGAISYVN 360
Qy 361 LSAGMSNFLAPALVLLPWFSTIGVYATYATLALFVLCALFIRVEDPGSSAPVTEKA 420
Db 361 LSAGMSNFLAPALVLLPWFSTIGVYATYATLALFVLCALFIRVEDPGSSAPVTEKA 420
Qy 421 LNIS 424
Db 421 LNIS 424

RESULT 2
US-09-738-626-3627
; Sequence 3627, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3627
; LENGTH: 445
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3627

Query Match 26.3%; Score 580.5; DB 9; Length 445;
Best Local Similarity 30.4%; Pred. No. 4.0e-42;
Matches 131; Conservative 89; Mismatches 188; Indels 23; Gaps 8;

Qy 9 LGPLHLHMGVIAIAVMTGDFELAFSLHYIKS-LGFTPAEASFATLYGLAALSAWV 67
Db 12 LGIRPLIFGFTIGTIFMIGDGETNLEPLSEHSGSVSLACTIVYVAVAAIAAF 71
Qy 68 SGVAAEITTPQKMLIGFVLCVHVLFVGLGQANGILLFYGIRGLAYPLFYSFI 127
Db 72 AAALSDLMGPKRWILGSIWIVFELFVLTALTDHTMLIFLAYGLRGFGYPRFAYGFL 131
Qy 128 VVILHNVRSESSSALGMYMAVSVGIVAGSYIPSTIPIRMG--EMQTLMLALAFCGAG 185
Db 132 VVIATATSPKQOLGIVGQFVFAFSGLPTLGAALVATISMQVNLTFEYETLLWVSLVAVIG 191
Qy 186 GVIAIMISLRHKTPTGHHNHLPRKFAELSRVATLTYNRIPLSSIVRIINTLSLFGFA 245
Db 192 SLIALLGVKE-RRGRNPLVANPDDVKOTLGQFKLLRNDRAARVYTIKTINSIPTYAMA 250
Qy 246 VIMPMEFVDELGTTSENIQVMAAFETTTIFS-----NIFGVIYAEKKGMKRVIRMECL 300
Db 251 VFFSFTDLDL-----KQWLSMFLITTYIAVNLPRNPFSGDRGMARVTFWGGSI 305
Qy 301 GMAASSLAFFYMPQYFGHNWMMAMIPALIGCTFVAAPVMAAVPALPEPKH 352
```

```
Db 306 GGAATLALVYFIIP-FGVQAGMSGVVGFTITAGALFVSLAGVPLSLATVSLDPRHP 364
Qy 353 GAATSVNLSAGMSNFLAPALVLLPWFSTIGVYATYATLALFVLCALFIRVEDPGFS 412
Db 365 GAAMATYMLGVGAVAVPRLVANVPHPLIGPTGLIVMALYLLSGMWTLDLRTGTQRFOD 424
Qy 413 SAPYTEKALNT 423
Db 425 GVPALAEADAH 435

RESULT 3
US-09-738-626-6943
; Sequence 6943, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6943
; LENGTH: 595
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6943

Query Match 6.5%; Score 143.5; DB 9; Length 595;
Best Local Similarity 20.5%; Pred. No. 0.00014;
Matches 105; Conservative 70; Mismatches 154; Indels 183; Gaps 26;

Qy 20 IATAVMTGDFELAFSLHYIKS-LGFTPAEASFATLYGLAALSAWVSGVAEITTPQK 79
Db 69 IVALTVWGIGFKDSFTNFASSALSAAVDNLMGMAFIIFG-----TVFVFIVVIAASKFG 123
Qy 80 TMLIG-----FVLMCVFHVLFVGLGQANGILLFYGIRGLAYPLFYSFIYV- 129
Db 124 TIRLGRIDEAPEFRVSN-----ISMMPAGM--GIGLMFY--GTTPELFTFRNGVPG 172
Qy 130 -IHNVRSESSSALGWY--MAVSVGIVAGSY-----ISFTIPIRMGEMTL 175
Db 173 HDEHNVGVAAMSTTMEHNLHPMAIYAI-VGLAIAYSTFRVGRKQLLSAFPLGEGAE 231
Qy 176 -WL-----ALAF--CFAG-----GVIAIMISLRHV 196
Db 232 GMLCKLIDILAIITVFTAGSLGALQIGLSAANIIEDPSDMWTIGIVSLTLAFI 291
Qy 197 KTPGHHNHLPRKFAELS--RAVTLTYNRIPLSSIVRI----- 236
Db 292 -----FSALISGVKGIOYL-SNANVLAALLAIFVYVVGPTVSIINLPGS 336
Qy 237 --NTLSLFGAVIMPMEFVDELGTTSENIQVMAAFETTTIFS-NIFGVIYAEKKGMKRV 293
Db 337 IGNVLSNFPQAGRTAASADG---TAGEMLGSMITTFYAWAMWISNPFVGMFLAISRGRS 393
```


Db 309 SVFAIAFVTRNTLLISVLVLGTLGKIALDPIMVTVTKHAPKALGTTLSAIVFGMSG 368
QY 367 NEFLAPAIIVLLPWFSTIGVIAVYATYALYLAFLVLCAP 403
Db 369 SILAPVYTGVLADPAGSMQVGF-YLSCVLLVIGLLAF 404

RESULT 6

US-09-738-626-6555
; Sequence 6555, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKOTO
; APPLICANT: SENOH, AKIHITO
; APPLICANT: IEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; LENGTH: 449
; SEQ ID NO 6555
; TYPE: PRF
; ORGANISM: *Corynebacterium glutamicum*
US-09-738-626-6555

Query Match 6.3%; Score 138.5; DB 9; Length 449;
Best Local Similarity 21.0%; Pred. No. 0.00027;
Matches 88; Conservative 53; Mismatches 125; Indels 153; Gaps 21;

QY 7 QWLG--LPLHLIMGY---IAIAVMTGDGFELAFLSHYIKSLGFTPAE-ASFA----- 53
Db 127 EMEGMMLPVGFVGFGTAAGVGTALLESSGAEMAM-----SLGFTSATVGTFAAIVGCI 180
QY 54 FTLYGL-----AALASNAV-----SGVVAELITTPQKTMILGIVLMCVHVLFLVFGLGQAN 104
Db 181 FTTWGIKKGTAAAPQOLPMDLRSGYIDKLSDRPS-----IGKASTN 222
QY 105 YGLLLFVGIRGLAYPLFLYSFIVYIIHNVRSESSALGMYNAVYSGVIGVAGSYIPSF 164
Db 223 PSAL-----EPLALHTGIILLTVAV-----AYSINQWL-GSMFPTV 257
QY 165 TPIIMGEMGLMLALAFCFPAGVYIAMISLRHVKTGPHMHNLTTPREKFAEL-SRAVTTLLYTN 224
Db 258 QIDP-----FAMSFVVGIVGMGIMRLKKPEYL-----DRDTVNSVSGAAT----- 298
QY 225 RNIFLSSIVRIINTLSLFGFAVIMPMFVDELGFTTSEMIQWMAAFETTTIFSIIFGCIY 284
Db 299 -----DYLIAFGIASTIAPAIAID-----YMWPLVLVFLVGLTYICCFE--- 336
QY 285 AERKMWRVIRWFGCLGMAASSLAFFYMPQYFGHNYWMAMIPALAGTFVAAPFPMNAV 344
Db 337 -----FWVAPRFEGEKMLERAI--FGKGMATAAATGIAL 370
QY 345 PALEPKHKAISVYNLS-----AGMSNFLAPAIAYVL-----LPWFSTIGVIAV 390
Db 371 KIYDPKLSGALNBYGVAAYIGAPFEIGMT-IAP-IAYLAGFTMGSMASLIAYIYIF 427

RESULT 7

US-09-815-242-10467
; Sequence 10467, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815, 242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10467
; LENGTH: 425
; TYPE: PRF
; ORGANISM: *Escherichia coli*
US-09-815-242-10467

Query Match 6.2%; Score 136.5; DB 10; Length 425;
Best Local Similarity 20.3%; Pred. No. 0.00038;
Matches 93; Conservative 63; Mismatches 136; Indels 161; Gaps 23;

QY 8 WLGLPLHLIMGYIAIAVMTGDGFE---LAFLSHYIKS-LGFTPAEASFAETTYGLAAL 63
Db 41 WLGY-----VE---DGFDFMIFYLIIHIIKADGIDPDIQATLIGTAVFIARPI 85
QY 64 SAWSVGYVAELITTPQKTMILGIVLMCVHVLFLVFGLGQANTGLLLFYGIRGLA---YP 120
Db 86 GCGEFGAMADKY-GRKPM-----HMAIF-----IYVGT-----GLSGIATNLVM 125
QY 121 LFLYSFIVYI-----IHNVRSESSALGMYNAVYSGVIGVAGSYIP----- 162
Db 126 LAYCRFLVIGLMSGEVYACASTYAVESPKNIQSAFSLVSGSVGIIIAQIIPQAEV 165
QY 163 -----SFTPIIMGEMGLMLALAFCFAGVYIAMISLRHVKTGPHMHNLTTPREKFAEL 217
Db 186 YGWRNSPFIGLPLVILVMI-----RKSAPESQEW 215
QY 218 VTLTYTRNIFLS-----SLVRIINTLSLFG-----FAVIM 249
Db 216 IEDKYKDSFSLSVFRKPHLSISMIYELVCFCLFGAMPINGLLPSYLAONGVNTVVIS 275
QY 250 IMFVDELGFTTSEMIQWMAAFETTTIFSIIFGCIYAEKGMWRVIRWFGCLGMAASSLA 309
Db 276 LMTIAGC-----TLNGTTFEGVGDIGYKKAFF-----VGLITSTI-- 313
QY 310 YMPQYFGHNYWMAMIPALAGTF-----VAAPFPMNAVPALEPKHKAIS-VYNLSA 363
Db 314 FLCPLEFTISVKNSSLIGLCLFGLMFTNLGIAGLVP-KFIYDPTPKLRGLGTGLIYNLGA 372
QY 364 --GMSNFLAPAIAYVLVLPF--STIGVIAVYATYALYL 396

[illegible]

```

Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6196
LENGTH: 701
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6196

Query Match          5 98: Score 130; DB 9; Length 701;
Best Local Similarity 22.38; Pred. No. 0.0024;
Matches 91; Conservative 60; Mismatches 158; Indels 99; Gaps 20;

Oy 51 SPFAFLYGLAALASAMVSGV-----VAELTPQKTMLGCVLHVLDFVGL----- 100
Db 154 AIAFEQVALARAM-AWAGFAFTIGIAGLC--RTWLQPLFLFAGSIITITVPLGLEGSHA 210
Oy 101 --GQANYG-----LILFTYGRGLAVPLFLYSFTVIINHVR-SENSSALGWY--WA 148
Db 211 AGGDHDYGTNSLHMLVLMILMWGL-----MALIAHARRIGPMNDMAVKKYSIIA 261
Oy 149 YVSV-GIVAGSYIPIFTPIPMGEKGTLMALAFCEAGC--YIAMSIRHYKTPGHMNL 205
Db 262 TYAVIGMAISGVNALLRMDPSDLFTTYNGLLVFKAAGVVVVGMEGLAH-----RTFTI 316
Oy 206 TPREFAEISRAVTLTYNNRNIETSSIVRIINTLSLGFAYIM-----PMMEVDELGETT 260
Db 317 PKLENDPRNSALFTTI-----AIYEVLMAAVLTGVAISMGTTPRPAPRIQDLSVNA 367
Oy 261 SEW-----LCQVMAFEFTTIFSNIT-----FWGIV-----AEKMGMRVIRW- 296
Db 368 LEMGSLKEPFLINLVFTMMRFDLMIGTIGIMLAFYMGGLVALBRQAKKMMHRTFWVV 427
Oy 297 FGC--LGMAASLATTYYPQYFGHHYMMAMIPALILGTFFVAFFVMAVFRALDEP----- 349
Db 428 LGCITLVVTVSSGIGMNPATFTSMRIVAHMLISMVVPFVLVYAGPLSLIMEVABGEPR 487
Oy 350 --KHKGASISVYNLSAGMSNFLAPALAVVLLPWFSTIGVIVATYLL 395
Db 488 PGLHMAAVMTDN-----PLKFTIMHPAVNTIOFTITFVALYL 525

RESULT 10
US-09-738-626-6877
Sequence 6877, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO

```

APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MAKOTO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6877
LENGTH: 453
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6877

Query Match 5.98; Score 129.5; DB 9; Length 453;
Best Local Similarity 21.38; Pred. No. 0.0016;
Matches 97; Conservative 75; Mismatches 180; Indels 103; Gaps 22;

QY 33 LAFSHYIKS---LGTTPAESFAFTLYGLAALSAWVSQVAE-IITPOKMLIGFVLM 88
DB 5 LAFLYSYVDGGLGMOTALSLVAGGQVYMTSLVASTADRVLSKENTLYSAIIV 64
QY 89 CVFHV-LFLVFGLOANGLILFYGRGLAVPLFLVSEIVYITHNVSSESSALGMYW 147
DB 65 MLGHIALALPGYGLSLGVLIGLIGSGGVK-----TAAQVVLGOLYSRTDTRDAGF- 117
QY 148 AVSVGIGVAGSYIPSEFTPIPMGEMGLM---LALACFAGGCVAMISLRH--VKTCGH- 201
DB 118 SIFMGVNLGGLFPLLTNALMGWGRHMGFGIA-AVGMAIGLQVYAMRTTIGAGHT 176
QY 202 MHNLTPEKFAE-----LSRAVTLTYNINFLSSIVRIINTLSLFGFVIMPMFVD 254
DB 177 VPNDLPKNEVARIIGAVVAVALIATGIITKLEMLSNITAIALIALAOMYVS 236
QY 255 E-----LGR-----TTSEMLQVMA-----APF-----F 272
DB 237 PLTTAAKSRLLGPIPMFVGVLFAIFOTFTVLAVYSDTRLDNFEGLDLPGLINSF 296
QY 273 TTFSNFMGIVA---EKMG---MMRVIRMGCLGMAASSLAFFYMPQYRGHN----- 319
DB 297 NPITIIIFSGIFATLMTKLGAKOMSTAVK-FGVANIVIGCALFEFLPAGAEENSTPMAL 355
QY 320 -YMMAMIPAIN-----LGTFAVAFVPMVAFPALPEKHKGAISV-YNLSAGMSN 367
DB 356 IIVVYFLETTIAELLSPVNSLATKVAPEAFQSRMFAVWL-----MAVSMGTSLSGTGLG 410
QY 368 FLAFAIVALLPMWESTIGVYATATLYLAFVLC 402
DB 411 YDPTDAGSEKVEFTIGV-----AAIVLGAIVIAA 441

RESULT 11
US-09-738-626-4815
Sequence 4815, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MAKOTO
APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4815
LENGTH: 450
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4815

Query Match 5.88; Score 128; DB 9; Length 450;
Best Local Similarity 21.28; Pred. No. 0.0022;
Matches 90; Conservative 62; Mismatches 163; Indels 110; Gaps 21;

QY 80 TMLIGEVLM-CVPHVLEVLG-0ANYGLILFYGRGLAVPL-FLYSFVYIHNVR 135
DB 42 SMIGFCVWLYVSAIADLNRIGFLDSAGOLYMLASIPGLAGLIRLIYELPLIGTRK 101
QY 136 SENSSAL-----GW-----YV-----AVSVGIGVAGSYIPS----- 163
DB 102 LVGISGGLFLPMFGLANVODSSTPYWMLTLALIGIGGVSGYMPSTGYFFPKAKS 161
QY 164 -FTPIPMGEMGLMIALAFCFAG-----GVYAMISLRHV-TEGHMHN---LTPR 208
DB 162 GTALGIDAGIGNLVSI-IDPMGPVMVGFGLGIGFLTPORTIGCTVVFVMAIYLVPM 220
QY 209 EKPELSRAVTL-----LYTNINFLSSIVRIINTLSLFGFVIMPMFVD 254
DB 221 TILAAVLSFLKDPVYANRQOIIDFGNKNITLSIYIMTGPAGPAGAOGLINN 280
QY 255 ELGFTT-----SEMLQVMAFF-----TTFSNFMGIVAEMKG---MMRVIRMGCL 300
DB 281 NFGIASPMAETYPADMILHAGTFAFLGPLICALYRAMGPLCDFFGCAINTFV---GCI 336
QY 301 GM--AASSLAFFYMPQYFGHYYMMAMIPAIN-----GTFVAFVPMVAFPALPEP 349
DB 337 GMTATATAAATFISRAETPDDEFFELMSMLALFFFTGLGNAGTF-----KOMPILP 388
QY 350 KHKGAISVYMLSGMSNFLAPATAVLLPWFSTI-----GVYATATLYLAFVLC 405
DB 389 KROAGGVIGW---TGAIGARFPGFIVGVLLSFTPTVAFEFMGCVFIIATATLTIYVARN 445
QY 406 VEOPG 410
DB 446 APFPG 450

RESULT 12
US-09-922-501-10
Sequence 10, Application US/09922501
Patent No. US20020120119A1
GENERAL INFORMATION:
APPLICANT: DARTOIS, Veronique A.
APPLICANT: HOCH, James A.
APPLICANT: VALLE, Fernando
APPLICANT: KUMAR, Manoj
TITLE OF INVENTION: 2,5-DKG PERMEASES
FILE REFERENCE: P-SR 4877
CURRENT APPLICATION NUMBER: US/09/922,501
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US/09/633,294
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US/09/677,032

;; PRIOR FILING DATE: 2000-09-29
;; NUMBER OF SEQ ID NOS: 22
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 10
;; LENGTH: 429
;; TYPE: PRT
;; ORGANISM: Pantoea citrea
US-09-922-501-10

Query Match
Best Local Similarity 20.8%; Pred. No. 0.003; DB 10; Length 429;
Matches 94; Conservative 77; Mismatches 170; Indels 110; Gaps 24;

QY 17 MGYIAVMTGDFELAFSLHYKSLGFTPAEASFATLYGLAALASAVGVAEI-1 75
DB 12 MWYLMVFIT---YSLAYDR-----ANYSFAA-----SGIEDLGI 47
QY 76 TPQKMLIGFVLMCFVHLELVFGLOANTGLILFYGIRGLAY-PLFLYSFIV--11 131
DB 48 SRGTSSLIG---ALFFLGFIYFQVPAIYA---VKRSVKLVFTSILLMGFCAAAATGLI 100
QY 132 HNYRSENSSALGMYAVYVSGIGVAGS-----YIPSF-----TIPINGEMG 173
DB 101 SNIPA-----LMVIRFVLGVBAVMPMLIYISNFTROBSRANTFLVLGNPV 150
QY 174 T-LMLALA-----FCFAGGVIAIMI-----SLRHVKTPGHMNTLPREK-- 210
DB 151 TVLAMSIVSGYLINARGMRFMEIFE-GVPALIMAFIMFWFIYR--DKPEQVSWLTEREKQ 207
QY 211 ----FELSAVATLL-----YTNRNIFLSSIVRIINTLSLFGFVAVIMPMF--VDLGF 258
DB 208 LASMAEEOAIPMRNVPOALRSRNVVYLCILHALMSIGVGFMMMPILRSASMDI 267
QY 259 TISEMLOVMAAFPTTIFSNIFMGIVAEKMGMRVIRWFCCLMAASSLAFYMPQFCH 318
DB 268 VRGGMIA--AVPYLAAILITMLVTSWLSDKTGLRRLFTPLLL--IASVFFESWMLGSI 322
QY 319 NYMMAM-IPAALGTVEVAFAVMAAFALEPKH-KGAISVYNLSAGNSNPLAPAIAYV 376
DB 323 SEFESYGLVLVAACWAPYGPFFALLPELPRNVAGISIGLINCCGALGAFGAMLVGY 382
QY 377 LFPWFSTIGVAVIYALYLLAFV-LCAFIRV 406
DB 383 LNLGTGGPGASVTYFMAIALLVSVGLVEFLKV 413

RESULT 13
US-09-815-242-11180
; Sequence 11180, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625

;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 11180
;; LENGTH: 407
;; TYPE: PRT
;; ORGANISM: Haemophilus influenzae
US-09-815-242-11180

Query Match
Best Local Similarity 20.8%; Pred. No. 0.003;
Matches 94; Conservative 71; Mismatches 161; Indels 125; Gaps 27;

QY 4 NKKOMGLPLHLIMGYIAIAVMTGDFELAFSLHYKSLGFTPAEASFATLYGL 59
DB 6 NSYGMKAL-----IGSAVGYGMDGFDLLILGFMLSAISADLNTLPAOGGSLVWTLI 57
QY 60 AALASAVGVAEIITPQKMLIGFVLMCFVHLELVF-----GLGOANYGLILFYGR 115
DB 58 GAFFGGLFGLALSD-----KYGRVRLVWTI--LFAVFTGLCAIAO-GTWDLLITRTIA 109
QY 116 GL-----AYPLFLYSFIVYIHNVRSENSSALGMYAVYVSGIGVAG-S 159
DB 110 GIGLGEFPGMALAEAWP-----ARHAKAA-SYVALGMO-----VGVLGAA 152
QY 160 YIPSETIPINGEMGTLMALAFCFPAGGVIAIMISLRHVKTGPHMN-----LTPRE 209
DB 153 LTPLLPLPHIGMGMLVYGIFPAFVAMFLR-----SHLDEPEIFQKOTALSTOS 202
QY 210 KEAELSAVATLLYTNR--NIFLSIV-RIINTLSLFGFVAVIMPMFVDELGTSEMD 265
DB 203 SEFDKLRSPOLLTKDRATSKISIGIYVLSVONFGYIGIMVLPNLSKOLGFSLTR-SG 261
QY 266 VMAAFFETTFESNIF-MGYIAEKMGMRVIRWFCCLMAASSLAFYMPQFCHNYMMAM 324
DB 262 LMTAVVCGMAGIWIIFGOLADRIKRPFLFO-LGAVSIYVYQSLTD----- 310
QY 325 IPAIAL-GTFVAFVP-----MAAVPALEPKHKAISV-YNSAGNSNPLAPA 372
DB 311 -PDIMLAGAFLMGFVNGMLGCGALMAEAYPR---EARATQONVLFNIGRAVGF-GPV 365
QY 373 IAVVLLPWFSTIG-VYIAY---TALYLLAFV 399
DB 366 V-----VGSVVAIVSFOFALIALAI 386

RESULT 14
US-09-738-626-6795
; Sequence 6795, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAKO
;; APPLICANT: SENOH, AKIHRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; PRIOR FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6795
LENGTH: 448
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6795

Query Match 5.6%; Score 124.5; DB 9; Length 448;
Best Local Similarity 22.2%; Pred. No. 0.0043;
Matches 102; Conservative 66; Mismatches 163; Indels 127; Gaps 23;

QY 29 DGEFLAFSLHYIKSL-----GFTPAEASFATLYGLAALSAWSGVAAEITTPQKTMILG 84
DB 41 DAMVGLISFPMALALAHWGSLSPETSLGSGFVGMAIGSLGGLADKIGRRQ----- 95
QY 85 FVLMCVFHVLFVYGLG-----QANYGLILFYGIRGLAY-----PLFL 123
DB 96 -----VFLSLVYGVATGASALSLSLMLALRFVVGGLGAEPLVASTLISESPRKY 150
QY 124 YSFIVYIIHNVRSESSSAAGMWAVYVSGIGVAGSYIPSTTIPIMGEMGTMLALACF 183
DB 151 RGRNVVL-----EAFMALGIMAAIVGTFFVAGS-----DNGMRM-ALALGC 192
QY 184 AGGVYAMI-----SLRHVKTPEGHMHN-----LTPREKFAELSRAY----- 218
DB 193 VPALVAVYVRLGLPESVRFLEKKGR-HDEAAIVVSFEEMAAAGKADATTAVIDHNA 251
QY 219 -----TLTYNNIFLSIVRIINTLSLGFVAVIMPMFVDELGFTTSMLOVMAAF 271
DB 252 EGSVYSMAALRKRTVALMWYPCIN-LSYAGAFIWIPSLVAD-GFTLVK-----SFQ 303
QY 272 FTTI-----FSNIFMGIVAEKMGWVRVIRMGCLGMASSLSLAFFYMPQYFGHNYMA 323
DB 304 FTLLITLTAOLPGYAVAAH--LIEKMGRRSTLATF--LVGSAISALYGL-----ANVEQI 355
QY 324 MIPALALGTF-VAAFPMAAVPALPEKH-----KGAISVYNLSAGNSNLAFAIAVYL 377
DB 356 LVAGCLSLFENLGAAGALYALGPELPTNVNGTGTGAA-----AGFGR-IASIIAPLI 407
QY 378 LPMFSTIGVVIATYALYLAF---VLCAFTVEQPGSSA 414
DB 408 VPPVIAFGPIALFALFATAPAFIAIAAFTLLPEQKGRSLA 447

RESULT 15

US-10-024-623-29
Sequence 29, Application US/10024623
Publication No. US20020187524A1
GENERAL INFORMATION:
APPLICANT: Curtis, Rory A.J.
TITLE OF INVENTION: 8099, 46455, 54414, 53763, 67076, 67102, 44181,
TITLE OF INVENTION: 67084PEL, AND 67084 ALT, HUMAN PROTEINS AND METHODS OF
TITLE OF INVENTION: USE THEREOF
FILE REFERENCE: MNI-214CP
CURRENT APPLICATION NUMBER: US/10/024,623
CURRENT FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/256,240
PRIOR FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: US 60/256,588
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: US 60/258,028
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 472
TYPE: PRT
ORGANISM: Escherichia coli
US-10-024-623-29

Query Match 5.6%; Score 123.5; DB 9; Length 472;
Best Local Similarity 20.8%; Pred. No. 0.0055;
Matches 98; Conservative 73; Mismatches 163; Indels 137; Gaps 27;

QY 19 YIAIAVMTDGDGEFLAFSLHYIKSLGFTPAEASFATLYGLAALSAWSGVAAEITTPQ 78
DB 24 FVSVAAVAGLLEGL-----DIGVAGALPEITDHFVLTSLQDEW-----V 65
QY 79 KTMILGFVLMCVFHVLFVYGLGOANY-----GLILFYGIRGLAYPLFYSFVVIHNV 134
DB 66 SSMMLGAAIGALFNG-WLSFRLGR-KYSLMAGALLFVLGSGAFATSVEMLIARV--- 120
QY 135 RSENSSALGMYMAVYVSGIGVAGSYIPSTTIP-NGEMG-----TLML 177
DB 121 -----VLGIAVGA-----SYTAPLYLSEMASENVRGKMISMVQLMVLGI 161
QY 178 AL-----ARCFAG-----GVIMISL-----RHVKTPEGHMHNL 205
DB 162 VLAFSLDPAFSYSGNMRRAMIGVLLAPVLLIILVFLPNSBRMLAEKGRHIEAEVLRML 221
QY 206 -----TPREKFAELSRAY-----TLTYNNR-----IFLSIVRIINTLSLGFVAVIM- 248
DB 222 RDTSEKAREELNEIRESLKTKLGSGMALFKINRNVRAVFLGMLLQANQFT--GMNIIY 279
QY 249 --PMFVDELGFTTSE-----WLOVMAAFETTTIFSINFCIVAEKMGWVRVIRMGCLG 301
DB 280 YAPRIF-KMAGFTTTEOQMIATLVVGLTFMEAT-FIVAF---TVDKAGRRPAK-IGFSV 333
QY 302 MAASSLAFFYMPQYFEGHN-----YNNAM-IPALALGTFVAAFPMAAVF-PALEP-KHK 352
DB 334 MALGTLVLGICLMOFDNGTASSGSLWSLGVGHTMNCINGYMSAAPVMIICSEIPLCKR 393
QY 353 GAASVYNLSAGNSNLAFAIAVLLPFSTIGVVIATYALYLAF 403
DB 394 DFGITCGTTTMMVSNMIIIGATFLLDSIGAAGFEMLYTALN-IAFVGITF 443

Search completed: March 13, 2003, 16:51:26
Job time: 14.9338 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:50:47 ; Search time 14.7417 seconds
(without alignments)
2765.009 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211

Sequence: 1 MSRRNQWMLGPLHLIMGYL.....RVEDPGSSAPYTEKALNIS 424

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1917	86.7	427	2	S78599
2	821	37.1	435	2	H69607
3	709.5	32.1	414	2	D68895
4	181	8.2	462	2	D83117
5	180.5	8.2	456	2	A71501
6	176	8.0	456	2	A10384
7	174	7.9	388	2	A37838
8	173.5	7.8	444	2	G70042
9	168.5	7.6	457	2	E71614
10	164.5	7.4	386	2	G64494
11	162.5	7.3	426	2	AF0634
12	161.5	7.3	400	2	C70082
13	159.5	7.2	428	2	AD0675
14	158.5	7.2	387	2	C96640
15	158.5	7.2	455	2	F86573
16	158.5	7.2	456	2	A72050
17	158.5	7.2	455	2	G81660
18	155.5	7.0	419	2	A83133
19	155	7.0	412	2	E84208
20	154.5	7.0	380	2	S74023
21	153	6.9	405	2	T44249
22	153	6.9	405	2	AE0990
23	152	6.9	396	2	F69813
24	151.5	6.9	428	2	B70078
25	151	6.8	413	2	B84085
26	150.5	6.8	409	2	D75358
27	150	6.8	401	2	AH1584
28	150	6.8	437	2	H83175
29	149.5	6.8	425	2	A90055

30	149.5	6.8	567	2	C75340	probable L-lactate
31	148.5	6.7	402	2	A83398	probable MFS trans
32	147	6.6	401	2	A81231	antibiotic resista
33	147	6.6	431	2	D81282	probable efflux pr
34	146.5	6.6	436	2	G72299	conserved hypotet
35	145.5	6.6	381	2	F71078	hypothetical prote
36	144.5	6.5	378	2	A99263	transpore membrane
37	144	6.5	429	2	A83574	multidrug resistan
38	143.5	6.5	420	2	AC0958	probable membrane
39	143	6.5	456	1	C64772	probable transpore
40	143	6.5	743	2	AE3145	hypothetical prote
41	143	6.5	747	2	F98142	iron(III) ABC tran
42	142.5	6.4	413	2	B87451	membrane protein,
43	142.5	6.4	454	2	AD0556	hypothetical major
44	142	6.4	406	2	AH0563	fornidomycin resis
45	142	6.4	463	1	A35620	peptide transport

ALIGNMENTS

RESULT 1

S78599
probable D-ribulose transporter - Klebsiella pneumoniae
N:Alternate names: ribitol transporter
C:Species: Klebsiella pneumoniae
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #ext_change 08-Oct-1999
C:Accession: S78599
R:Henel, H.; Turgut, S.; Schmid, K.; Langelier, J.W.
J. Bacteriol. 179, 6014-6019, 1997
A:Title: Substrate recognition domains as revealed by active hybrids between the D-ar
A:Reference number: S78599; MUID:97464425; PMID:9324246
A:Accession: S78599
A:Molecule type: DNA
A:Residues: 1-427 <HEU>
A:Cross-references: EMBL:AF045244; EMBL:U07127; NID:92905642; PIDN:AAC26496.1; PID:92
A:Experimental source: strain KAY2026
C:Genetics:

A:Gene: rbtR

Query Match	86.7%	Score 1917;	DB 2;	Length 427;
Best Local Similarity	85.6%	Pred. No. 1.2e-134;		
Matches 363;	Conservative 28;	Mismatches 29;	Indels 4;	Gaps 1;
OY	1	MSRRNQWMLGPLHLIMGYIAVEMTGDFELAFSHYTKSGTPPAKSPAFLYGLA	60	probable L-lactate
DB	1	MSVNNQWMLGPLHLIMGYIAVEMTGDFELAFSHYTKSGTPPAKSPAFLYGLA	60	probable MFS trans
OY	61	AALSANVSGVVAEITTPORTMLIGFVLMCVFHVLFVFGIGQANVGLILFFGIRGLAPP	120	antibiotic resista
DB	61	AALSANVSGVVAEITTPORTMLIGFVLMCVFHVLFVFGIGQANVGLILFFGIRGLAPP	120	probable efflux pr
OY	121	LPLYSFIVYIHNVRSESSSALGWYAVYSGVIGAGSYTSPFTPIPGEMKTLALAA	180	conserved hypotet
DB	121	LPLYSFIVYIHNVRSESSSALGWYAVYSGVIGAGSYTSPFTPIPGEMKTLALAA	180	hypothetical prote
OY	181	FCGAGGVIMISLRHYKTGCHMNLTPREKFAELSAVTLTYNNRNLPSISYRIINTLS	240	transpore membrane
DB	181	FCGTGGIIVLSMRHETPRHMONLTTRKFAELGRAATLLYNNRNLPSISYRIINTLS	240	multidrug resistan
OY	241	LEGFAVIMPMVEDELGTTSSEMLQVMAFFFTTIFSNIFMGVIAKMGMRVIRMGCL	300	probable membrane
DB	241	LEGFAVIMPMVEDELGTTSSEMLQVMAFFFTTIFSNIFMGVIAKMGMRVIRMGCL	300	probable transpore
OY	301	GMASSLAFLYVDFQHGHNFMALVPALALGIFVAFVPMAAVFPALDFPHKKAALSYVN	360	iron(III) ABC tran
DB	301	GMASSLAFLYVDFQHGHNFMALVPALALGIFVAFVPMAAVFPALDFPHKKAALSYVN	360	membrane protein,
OY	361	LSAGNENFLAPATAVLLPWFSTIGVIAVTAIYLAFVLCARIRVEQGEFS----	416	hypothetical major
DB	361	LSAGNENFLAPATAVLLPWFSTIGVIAVTAIYLAFVLCARIRVEQGEFS----	416	fornidomycin resis
OY	417	TEKA 420		peptide transport

..
 ..
 —
 —
 ..
 ..
 ..
 ..
 —
 —
 ..
 ..
 ..
 —
 —
 —
 ..
 —
 —
 —
 —
 ..
 —
 —
 —
 —
 ..
 —
 —
 —
 —

QY 235 IINTLSLFGFAVIMPMPFVDELGTTTSEWLQVWAAFFTTIFS----- 277

QY 235 IINTLSLFGFAVIMPMPFVDELGTTTSEWLQVWAAFFFTTIFS----- 277

Db 225 CILHLLSSSF-VALPOMMAN-AGLAPA---OHMVYLVLTMLVSEFAAVVFTIYAEMKRRM 279
OY 278 -NIMGVIAEKMGMVRVIRMFEGCLMAASSLAIFYMPOYGHNVMMAIRPALCTFPA 336
Db 280 KQVPEGCA-VLFAEVVLMFA-----GQDLMT-ITACVQI--FEIA 317
OY 337 FVPAVAEPALPEK-----HGGAISVYNLSAGNSNFIAPAIVAVLPM-FSTIGVITAY 390
Db 318 FNVWEALPILSLIKSESPAGYKGTAMGITS-----TSQIFGVAIGSLGMIIGLEGADNVF 373
OY 391 TALYLAFVLCAF-IRVEQPGFSSA 414
Db 374 AAGAIIALMFAVSVTQMPEPPYVSS 398

RESULT 7
A37838
nora protein - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 20-Jun-2000
C:Accession: A37838
R:Yoshida, H.; Bogaki, M.; Nakamura, S.; Ubukata, K.; Konno, M.
J. Bacteriol. 172, 6942-6949, 1990
A:Title: Nucleotide sequence and characterization of the Staphylococcus aureus nora gene
A:Reference number: A37838; MUID:91072245; PMID:2174864
A:Accession: A37838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <YOS>
A:Cross-references: GB:D90119; NID:g216974; PIDN:BAAI4147.1; PID:g216975
C:Superfamily: tetracycline resistance protein

Query Match 7.9%; Score 174; DB 2; Length 388;
Best Local Similarity 20.5%; Pred. No. 1,1e-05;
Matches 85; Conservative 76; Mismatches 191; Indels 62; Gaps 15;

OY 5 NKQMLGLPLHLINGYIAIAVMTGDFELAFLSHYKSLGTFPAEASFATLYGLAALS 64
Db 2 NKQIFVL-----YFNIFLIFLGIGIYIPVLPYKDLGLTGSDDLGLVAAPALSQMII 54
OY 65 AMWGVAAEITTPQKTMIGFVLCVPHVLEFLVGLGQANGLLPLFGIGLAPFLY 124
Db 55 SPFGTLADKLGKRLICIGLILFSVEFMFAV---GHNVSVLMSRVIGMSAGMMP 110
OY 125 SFVIVIIHNVSESSSALGMYWAVSYGIGVAGSYIPFTPIPMGEKGTMLALACFA 184
Db 111 GVTGLINDISPHQKAKNFGYMSAIIINSGF-ILGPGIGFMAEVSHRN-----PRTFA 162
OY 185 G--GVYAMIS---LRH---VKTGCHMNLPRKFAELSRVATLLYTNRIELSSIV-RI 235
Db 163 GALDILAFIMISIVLIHPKPKSTSGFQLEPO-----LLTKIMWKVFIPVITL 212
OY 236 INTLSLGEFAVIMPMFDELGTTSSE-WLQVMAAFETFTIESNIFMGIVAEKGMWVI 294
Db 213 VLSSGLSFEFLYSLYTRADKVNYSKPDSTAITGGGIFGALFOYTFD-----KFM 263
OY 295 RMFGCLGAASSLAF---YVMPQYFGHNYMMAMIPALGTFVAFAVMAV---FPAL 348
Db 264 KYFSELFIAMSLYSVVLILVAFANGWSIML--ISFVFIFGDMIRAPITVYFSIA 321
OY 349 PKHGAISVYNLSAGNSNFIAPAIVAVL-----PMESTIGVIAVATLYL 366
Db 322 GERGFAGGINSFTPTSMGNFIPLIAGALFDVHTAPYMAIGVSLAGVIVLI 375

RESULT 8
G70042
multidrug-efflux transporter homolog yvka - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G70042
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Bertet

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Exlington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gal
lechi, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koelter, P.; Koningsstein, G.; Krogh, S.; Kuno, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schreuter, R.; Scifone, F.; Sekiguchi, J.; Sekovska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G70042
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-444 <KUN>
A:Cross-references: GB:299122; GB:AL009126; NID:g2636029; PIDN:CABI538.1; PID:g26360
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvka
C:Superfamily: fosmidmycin resistance protein

Query Match 7.8%; Score 173.5; DB 2; Length 444;
Best Local Similarity 20.1%; Pred. No. 1,4e-05;
Matches 91; Conservative 87; Mismatches 185; Indels 89; Gaps 19;

OY 4 NKQMLGLPLHLINGYIAIAVMTGDFELAFLSHYKSLGTFPAEASFATLYGLAAL 62
Db 3 SRKEM-ALYSLALGAILVPIINSTIVAVLSISHTNENI---ASITVAVVYLYVMA 57
OY 63 LSAMWGVAAEITTPQKTMIGFVLCVPHVLEFLVGLGQANGLLPLFGIGLAPFLY 119
Db 58 VTQIAGKLGDMY-GNKTW---YLMGV--GLPIASLGALSPSLILLIFRLQAVCG 110
OY 120 PLFLYSFVIVIIHNVSESSSALGMYWAVSYGIGVAGSYIPFTPIPMGEKGTMLA 178
Db 111 ALTPNSIAIRHVVSERKLPKVPFGFG--LGAGLGAALGPIGSLILDSFMSHISRWV 168
OY 179 LAF-----CFAGVYAMISLRVVKP-GH----- 201
Db 169 IPIALALFTALMFPQYKENSADPDIIGSLILAGSIVSILLTNEAPWGTIVSVL 228
OY 202 ---MNLTPREKPAE-----LSRAVTLTYTNRIELSSIVRIITLSLGEFAVIMPM 250
Db 229 ILFLVPLFPRREKTPQPIIDFALFKST--FTNANLSV-----LSNLMKMYAVLIMPL 281
OY 251 MEVDELGFTTSEWLVMAAFETFTIESNIFMGIVAEKGMWVRVIRMFEGCLMAASSLAIFY 310
Db 282 FMTQFGLNLSNSGMAISVESIFMSASNMVGAQQLHHKMGAKKIT-FLSFYMMAGANLFL 340
OY 311 YMPQYFGHNYMMAMIRPAL-ALGTFVAFAVMAVFPALPEKHKHGAISVYNLSAGNSNF 366
Db 341 LLSS--SHSVLEFLSLILGLGSLGVLTSQVSLATVPDGMGSAVSIFSTRFGST 398
OY 369 LAPAIVAVLPMFESTIGVIAVATLYLAVL 400
Db 399 ISSAL-----IGLISGYHFLFMILFAV 420

RESULT 9
E71614
membrane transporter PFB0465C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71614
R:Gardner, M.J.; Tetteh, H.; Carucci, D.D.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pereira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9604551

A:Accession: E71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-457 <GAR>
A:Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AACT1882.1; PID:g384519
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0465c

Query Match 7.6%: Score 168.5; DB 2; Length 457;
Best Local Similarity 19.8%: Pred. No. 3.3e-05;

Matches 90; Conservative 78; Mismatches 154; Indels 133; Gaps 18;

OY 35 FLSHYIKSLGTFPA---EASFAETLYGLAAASAMVSGVAELITPQKMLIGFVLCVF 91
DB 38 YVSYMKIIGCSQVKKKSSMIYVLLTLQCFEGFGGILNQLQISVLGGWLMCG 97
OY 92 HVL-FLVGLGOANGLILFYGIR---GLATPLFLYSFIVIIHNVRSSSSALG 144
DB 98 ILLSYTFV---NEYLELMTYGLIGIGCGIAYPIPL-----SVAVKKHVDYKG 143
OY 145 WYMAVYVSGIGVA-----GSYIPSF-----TIPIMGEMGTML 177
DB 144 VSGIIFIGRGLSVFIFICPLQNTYINKYMPDYPEIENSDEKYSNLDILKKVPLFI 203
OY 178 ALAFCEAGGVAMIASLRHVKTPGHMHLTPREKFAELSRVTLTYNNRNF-----LSS 231
DB 204 YEGICFA--IIQFGLSYLADSGD---TSKDPMAVNDNRNKKVLFYEKKNFINKPGLSN 257
OY 232 IYRIINTLSLFGFAVIMPMFVDELGFTTSEMIQVNAAFFT-----273
DB 258 SLTTLSTNTSNFSREY-----NNTFINPEFLIMLIFFMQAISYTOVFKIFGMNY 310
OY 274 -----TFSNIFGIVAEKMGW-----MRVIMPGCLGMAASSLAF 309
DB 311 LSTDDBSLILGVSLSLNFICRIEFGILSDPFSKTTLLMSLMSF--LTTTLWSGF 368
OY 310 YVMPQYFGHNYMMAMIPALALGTFVAAFVMAVFPALF-----PKHGAISVYNLSAG 364
DB 369 YGIIIV---SIWCLIFECAGTF-----AIFPSITAHFGTINFGVFPOLLTFARA 417
OY 365 MSNFLAPALAVULLPMFSTIGVIAVYALYLAFAV 399
DB 418 FSSIIAMIIISAVLL--NNIGINMCAIVLSISFV 449

RESULT 10
G64494
quinolone resistance protein norA homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: G64494

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodex, A.;
ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Kleink, H.P.; Fraser, C.M.; Smith, H.O.; Moese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MID:96337999; PMID:8688087
A:Accession: G64494
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-386 <BUL>
A:Cross-references: GB:U67596; GB:L77117; NID:g2826430; PIDN:AA899579.1; PID:g1592184;
C:Genetics:
A:Map position: REV1536434-1535274
C:Superfamily: tetracycline resistance protein

Query Match 7.4%: Score 164.5; DB 2; Length 386;
Best Local Similarity 20.0%: Pred. No. 5.6e-05;
Matches 82; Conservative 91; Mismatches 189; Indels 47; Gaps 15;

OY 15 LHWGTAIAVPMTGDFELAFISHYIKSLGTFPAEASFAETLYGLAAASAMVSGVAELI 74

DB 10 VIV--ITFTTLMGVFIAPIMAIYAQTGATMLEIGLIFGSPALARTVAQIPVGLSDI 67
OY 75 ITPQKMLIGFVLMCVENHL-FLVGLGOANGLILFYGIRGLAVPLFLYSFIVIIH 132
DB 68 YGKKEFTVCGTFEYGVSTLMYNFVSVTLG--FLIVRIFGIFSAPVTPVAGSVIAIAP 124
OY 133 NVRENSSSALGWYMAVYVSGIGVAGSYIPSTPIPMGEMGTMLALAFCEAGGVAMIS 192
DB 125 KTR-----LGEYKILFNSAI--TLGFGICPFIIGLADMYGKMFYCGFGILAAI- 175
OY 193 LRHVKTTPGHMHLTPREKFAELSRVTL---LYNNRIFLSSIVRIINTL--SLFGFA 245
DB 176 ISYMKLEDIYFNKN-REKI-DVKKISTLSFEFLKRNFSSTIIVSNVMIAGIYAVL 233
OY 246 VIMPMF---VDELGFTTSEMIQVNAAFFTTIFSNIF--NQIVAEKMMKMMVIMFGC 299
DB 234 ALYAINYNTTSOVGFMA-----LTNTLMALORSFGKLYDKGNIMIT----- 278
OY 300 LGMMASSLAFFVMPQYFGHNYMMAMIPALALGTFVAAFVMAVFPALFEPKHGAISVY 359
DB 279 IGFIIFSGMYLLSTFTPLTILASLTIIVAGSSISSTATSLAVNDIPTRKGEAMGLF 338
OY 360 NLSAGSNFLAPALAVULLPMFSTIGVIAVYALYLAFAVLCAFIVEQ 408
DB 339 TTSINIGMFIQ-AVSEGLADIIGIANMYKFAISFIVGIIISYLRIR 386

RESULT 11

AF0634

probable transporter stry1169 [imported] - Salmonella enterica subsp. enterica serovar
C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typhi
C:date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-May-2002
C:Accession: AF0634

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
; S.; Moulie, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se
A:Reference number: AB0502; PMID:11677608
A:Accession: AF0634
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08257.1; PID:g16502304; GSPDB:GN00176
C:Genetics:
A:Gene: Stry1169
C:Superfamily: Escherichia coli hypothetical protein b4279

Query Match 7.3%: Score 162.5; DB 2; Length 426;
Best Local Similarity 20.6%: Pred. No. 8.7e-05;
Matches 89; Conservative 66; Mismatches 147; Indels 131; Gaps 17;

OY 59 LAALASAMVSGVAELITPQKMLI-----GVLMCVFVFLV---FGL-----100
DB 1 MIKKFPWIS---ELTRPQKALFSAWGLGYVDGDFMLFIYIMLIKRADLGLTMEGA 56
OY 101 -----GOANYGLILFYGIRGLAVPLFLYSFIVIIHNVRSSSSALGWYA 148
DB 57 FLATAAFIGRPGFAGLEGLADKFRK---PLMMSIVAYSVGTGLSGIASGVIMULTLS 112
OY 149 VYVSGIGVAGSY-----JFSFTPIPMGEMGTML 176
DB 113 RFTVGMGMAGKACASTYAVESWPKHLKSKAFLVSGFGIGNIITAYFSPFAE-AVGA 171
OY 177 LALAFCEAGGVAMISLR-----HVKTTPGHMHLTPREKFAELSLRA 217
DB 172 RAAFEGGLPVLVLIYIRARAPESKEMBEAKLSGLKHSASMSVSLSKGLF-----225
OY 218 VLLTYNNRIFLSSIVRIINTLSLFG---FAVIMPMFVDELGFTTSEMIQVNAAFFTT 273

Db 226 -----NRAQPLTLCVFIYLFSTFGANWPIEGLLPTYLAGE-GEFQGVSNLMTAAFG 278
Qy 274 TIFSNIFWGIWAERKGMWRVIRWFGCLGMAASLAFYMPQFGHNWMM-----AMIPAL 330
Db 279 TIVLNIWGLCADRIIGLKTFF-----SIGLMSFLFIPLFPIPDNDYLLGACLFGLMAT 334
Qy 331 GTFPAAYP-MAAVPALPEKHKAISVYNLSA-----GMSNFLAPAI 375
Db 335 NVGVCGLVPKFLYDFP-LEVRGLGTGL-IYNLAATSGTENSMAATWLGITMGLAALTFF 392
Qy 376 VLLPWFSTIGVI 388
Db 393 IVAFWTATILLI 405
RESULT 12
C70082
Multidrug-efflux transporter homolog yxhH - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_rev1501 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: C70082
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koether, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Meue
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, R.; Park, S.H.; Parro, V.; Pohl, T.M.; Portereid
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Satale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeder, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; M01D:98044033; PMID:9384377
A:Accession: C70082
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-400 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15890.1; PID:el186363;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yxhH
Query Match 7.3%; Score 161.5; DB 2; Length 400;
Best Local Similarity 21.5%; Pred. No. 9,7e-05;
Matches 93; Conservative 70; Mismatches 158; Indels 111; Gaps 18;
Qy 18 GYIAIVFMGQDGEFLAFLSHYIKSLGFTPAEASFATLGLAALASAWGVAAEII-T 76
Db 33 GYIVVPF-----SLYLEDLHFSYGAIGIILSGSYVTQIILRFLGLSLDILFS 81
Qy 77 POKTMLI-GFVLMGVFHYLFVFLGLGOANYGLL--LFVIRGLAVPL-ELTSFIYI 130
Db 82 LKROLLLAGFQSVLSSLLFLMPD---SEFFVLAARFAPITASTWMAATILVA----- 132
Qy 131 IHNRENSSSALG--WYMAVYS-----VGIGVAGSYIPSTFIPIMSEMGTLALAFCE 183
Db 133 -HYNNNGNASKAMKIMQFETVMPFASIVFCGLAAH-----LGRQVFFMALASA 183
Qy 184 AGVYIAMISLRHVTPGHMNLTPREKFAELSRVLT-LYTNRIIFLSIIRIINTLSLF 242
Db 184 AGVITCGFIQDPSAPRANRGIRVNOYIKDLRLPKLKT-----ILSMTHAVALFITVF 239
Qy 243 GFAVIMPMFVDELGFPTTSEMLOVMAAFETTTISNIFWGIWAERKGMWRVIRWFGCLG 302
Db 240 GFT-----PLYMNOGLMGDAELLMWSAF-----LPH 267
Qy 303 AASLAFYMPQFGHNWMMAMIPALGLTFVAFVMAA-----VF 344
Db 268 AAATLSFVFL-RTSRIAYSVMLISFAVTGVCLLVFPFLAALFTVCITHACIGALGFVF 326

Qy 345 PAL-----EPKHKAISVYNLSAGMSNFLAPAIAYVLLPWFSTIGVIAYATAYLL 396
Db 327 PLLSHVEISSARLKKMSVMQFYSGFALGFIPLLAGKIAQIIGLAGVFYGGASLAF 386
Qy 397 AFVLCAPIRVQ 408
Db 387 AF-----FVMLAQ 394
RESULT 13
AD0675
membrane transport protein STY1519 [imported] - Salmonella enterica subsp. enterica s
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:date: 09-Nov-2001 #sequence_rev1501 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD0675
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AD0675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01773.1; PID:g16502620; GSPDB:GN00176
C:Genetics:
A:Gene: STY1519
Query Match 7.2%; Score 159.5; DB 2; Length 428;
Best Local Similarity 20.3%; Pred. No. 0.00015;
Matches 99; Conservative 73; Mismatches 173; Indels 143; Gaps 19;
Qy 1 MSRRNKQMLGPIILMIGYIAIVFMG--DGEFLAFLSHYI-KSLGFTPAEASFATLY 57
Db 1 MKRRTGMGLV-----FLFIYMLNMYDRSALSTAPATELELGFNAEKGMTFSAF 53
Qy 58 GLAALASAWGVAAEITTPOKTMLIGFVLMGVFHYLFVFLGLGOANYGLLFPYGRGL 117
Db 54 FIGALNLFIDGMSADKVGPKTYFLAALLMSV-----FCGLTGL 93
Qy 118 AYPLEYSFIYV-----IHN-VRENSSSALGMYMAVYSVIGVAGSYI 161
Db 94 VTGLMTLIVLVLGMAEGPVSAGNKIINNMIKRESATVIGIFSAGSPVAGAVSG--- 150
Qy 162 PSFTIPIMGEGTLMALAPCFAGCVTAMISLRV-----KTPGHMNLTPREFAE 213
Db 151 ----PI--VGLLALSIGMPAPGIIIFLGLVWVLLMYFIVSDPKPTMSKRLAERID- 201
Qy 214 LSRVATLLYTNRIIFLSIYRIITLTLFGFAVYMPMFVDELGFPTTSEMLOVMAAFET 273
Db 202 -----FENNEDVILDDGKATPSL---GYMKQPMVMAATTLAFESYNI-----LFFFL 247
Qy 274 TIFS---NIWGIWAERKGMWRVIRW-FGCLGMAASLAFYMPQFGHNWMMAMIPAI 329
Db 248 TWFSYLNHSLDIDKEISLATVIRVYIGALGVLCVSDVIYRIGN--ALLSRLLI 304
Qy 330 LGTFVAAFVPMAAFRLPERKHKAISVYNLS----- 362
Db 305 LGVCLAGAAVCAVAVSGVST--IGSATITLMSVSLFLYLTGPIYMAVYDQVHDKVGSV 362
Qy 363 -----AGMSNFLAPAIAYVLLPWFSTIGVIAYATAYLLAFVLC-----AFIR 405
Db 363 GGAMHGLANISGIIIGPLVT-----GFIYQSGKYDAFYALAGALIVSSLLVVF 412
Qy 406 VEDQGFSS 413
Db 413 VKSKGFKA 420

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:42:27 ; Search time 8.42384 Seconds
(without alignments)
2087.642 Million cell updates/sec

Title: US-09-802-208b-5
Perfect score: 2211
Sequence: 1 MSRRNKQWLGPLHLIMGYI.....RVEQPGSSAPVTEKALNIS 424

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1917	86.7	427	1 RBT_KLEPN	052717 klebsiella
2	1907	86.3	425	1 DALY_KLEPN	052718 klebsiella
3	821	37.1	435	1 CSBX_BACSU	005390 bacillus su
4	709.5	32.1	414	1 YOAB_BACSU	034864 bacillus su
5	180.5	8.2	456	1 UHPT_CHLFR	084548 chlamydia t
6	174	7.9	388	1 NORA_STAM	P21191 staphylococ
7	158.5	7.2	455	1 UHPT_CHLFR	092709 chlamydia p
8	158.5	7.2	456	1 UHPT_CHLFR	092718 chlamydia m
9	158	7.1	413	1 MUOK_ACICA	P94131 acinetobact
10	151.5	6.9	428	1 YXIO_BACSU	P42306 bacillus su
11	144	6.5	458	1 TCR_STRAG	P13924 streptococc
12	143.5	6.5	595	1 BERP_CORGL	P54582 corynebacte
13	143	6.5	454	1 YABR_ECOLI	P77726 escherichia
14	141	6.4	497	1 DTPY_LACLA	P36574 lactococcus
15	140	6.3	429	1 G6PU_HUMAN	043826 homo sapien
16	138	6.2	428	1 BCH2_RHOCA	P26171 rhodobacter
17	137.5	6.2	400	1 BMR2_BACSU	P39843 bacillus su
18	137	6.2	388	1 Y535_BUCAL	P57601 buchnera ap
19	137	6.2	405	1 PMRA_LACLA	P58120 lactococcus
20	136.5	6.2	405	1 YUHB_ECOLI	P39352 escherichia
21	136.5	6.2	532	1 COX1_RHOCA	P98059 rhodobacter
22	135	6.1	463	1 PGTP_SALTY	P12681 salmonella
23	133.5	6.0	395	1 NARK_BACSU	P46907 bacillus su
24	133	6.0	419	1 YHHS_ECOLI	P37621 escherichia
25	133	6.0	458	1 EMRB_ECOLI	P27304 escherichia
26	132.5	6.0	458	1 TCR_BACST	P07561 bacillus st
27	132.5	6.0	458	1 TCR_STRAH	P36890 staphylococ
28	132.5	6.0	458	1 TCR_STRPN	P11063 streptococc
29	132	6.0	430	1 RFBX_SALTY	P26400 salmonella
30	131.5	5.9	389	1 BMR1_BACSU	P33449 bacillus su
31	131.5	5.9	473	1 PHDK_NOCCK	O24723 nocardioide
32	131.5	5.9	485	1 YJDL_ECOLI	P39276 escherichia
33	129.5	5.9	404	1 YYBF_BACSU	P37498 bacillus su

34	128.5	5.8	452	1 WZEY_SALTY	Q82396 salmonella
35	128.5	5.8	452	1 WZEY_SALTY	P37458 salmonella
36	128.5	5.8	459	1 TCR2_BACSU	P14512 bacillus su
37	127.5	5.8	433	1 TCR_STRAU	P02983 staphylococ
38	127	5.7	839	1 NAH3_DIDNA	Q28362 didelphis m
39	126.5	5.7	473	1 GLE_ZYMO	P21906 zymomonas m
40	126.5	5.7	515	1 VMT2_RAT	001827 rattus norv
41	126	5.7	422	1 EXUT_BACSU	034456 bacillus su
42	125.5	5.7	407	1 YB04_HAEIN	P71369 haemophilus
43	125	5.7	423	1 YDIN_ECOLI	P76198 escherichia
44	125	5.7	446	1 CITN_SALDU	P31603 salmonella
45	125	5.7	446	1 CITN_SALTY	P31604 salmonella

ALIGNMENTS

RESULT 1	ID	RBT_KLEPN	STANDARD:	PRT:	427 AA.
AC	052717				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Ribitol transporter.				
GN	RBT.				
OS	Klebsiella pneumoniae.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Klebsiella.				
OX	NCBI_TaxID=573;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=1033-5P14 / KAY2026;				
RX	MEDLINE=97464425; PubMed=9324246;				
RA	Heuel H., Turgut S., Schmid K., Lengeler J.W.;				
RT	"Substrate recognition domains as revealed by active hybrids between				
RT	the D-arabinitol and ribitol transporters from Klebsiella				
RT	pneumoniae.";				
RL	J. Bacteriol. 179:6014-6019(1997).				
CC	- SUBCELLULAR LOCATION: Integral membrane protein (potential).				
CC	- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX				
CC	SUBFAMILY.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL: AF045244; AAC26496.1;				
DR	InterPro: IPR004748; Ketoglu_permease.				
DR	InterPro: IPR003662; sub_transporter.				
DR	Pfam: PF00083; sugar_tr; 1.				
DR	TIGRFAMS: TIGR00897; 2A0118. 1.				
KW	Transport; Sugar transport; Transmembrane.				
FT	DOMAIN 1 7				
FT	TRANSMEM 8 28				
FT	DOMAIN 29 51				
FT	TRANSMEM 52 72				
FT	DOMAIN 73 79				
FT	TRANSMEM 80 100				
FT	DOMAIN 101 107				
FT	TRANSMEM 108 128				
FT	DOMAIN 129 141				
FT	TRANSMEM 142 162				
FT	DOMAIN 163 171				
FT	TRANSMEM 172 192				
FT	DOMAIN 193 238				
FT	TRANSMEM 239 259				
FT	DOMAIN 260 263				
FT	TRANSMEM 264 284				

```

FT DOMAIN 285 295 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 296 316 9 (POTENTIAL).
FT DOMAIN 317 323 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 344 10 (POTENTIAL).
FT DOMAIN 345 360 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 361 381 11 (POTENTIAL).
FT DOMAIN 382 383 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 384 404 12 (POTENTIAL).
FT DOMAIN 405 427 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 427 AA: 47009 MW: A9AB7EB9C26185AA CRC64:

Query Match 86.7%: Score 1917; DB 1; Length 427;
Best Local Similarity 85.6%: Pred. No. 1.7e-103;
Matches 363; Conservative 28; Mismatches 29; Indels 4; Gaps 1;

QY 1 MSRRNKQMLGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASFAFTLYGLA 60
  1 MSRRNKQMLGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASFAFTLYGLA 60
Db 1 MSRRNKQMLGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASFAFTLYGLA 60
QY 61 AALSAMWSGVVAELITTPQKTMILGCVFVHLFVFGQANYGILLFYGIRGLAYP 120
  61 AALSAMWSGVVAELITTPQKTMILGCVFVHLFVFGQANYGILLFYGIRGLAYP 120
Db 61 AALSAMWSGVVAELITTPQKTMILGCVFVHLFVFGQANYGILLFYGIRGLAYP 120
QY 121 LFLYSPFVLIHNVRSNSSAIGWMAVYSGVIGVAGSYIPSTTIPIMGEMGLMALA 180
  121 LFLYSPFVLIHNVRSNSSAIGWMAVYSGVIGVAGSYIPSTTIPIMGEMGLMALA 180
Db 121 LFLYSPFVLIHNVRSNSSAIGWMAVYSGVIGVAGSYIPSTTIPIMGEMGLMALA 180
QY 181 FCFAGVYIAMSIRHVKTPGHHNLTREKFAELSRATVLLTYRNIFLSSIVRIINTLS 240
  181 FCFAGVYIAMSIRHVKTPGHHNLTREKFAELSRATVLLTYRNIFLSSIVRIINTLS 240
Db 181 FCFAGVYIAMSIRHVKTPGHHNLTREKFAELSRATVLLTYRNIFLSSIVRIINTLS 240
QY 241 LFGFVAVTMPMFVDELFTTSEMLOVNAAPFTTIFSNIFGIVAERKGMKRVIRMGCL 300
  241 LFGFVAVTMPMFVDELFTTSEMLOVNAAPFTTIFSNIFGIVAERKGMKRVIRMGCL 300
Db 241 LFGFVAVTMPMFVDELFTTSEMLOVNAAPFTTIFSNIFGIVAERKGMKRVIRMGCL 300
QY 301 GMAASSIAFYMPQYEGHNWMAIPALGTFAAFVPAAPVPALEPKHKAISVYN 360
  301 GMAASSIAFYMPQYEGHNWMAIPALGTFAAFVPAAPVPALEPKHKAISVYN 360
Db 301 GMAASSIAFYMPQYEGHNWMAIPALGTFAAFVPAAPVPALEPKHKAISVYN 360
QY 361 LSGMNSFLAPALAVLLPFFSTIGVYIATYALYLAFLVCAPIFRVBPQSS----APV 416
  361 LSGMNSFLAPALAVLLPFFSTIGVYIATYALYLAFLVCAPIFRVBPQSS----APV 416
Db 361 LSGMNSFLAPALAVLLPFFSTIGVYIATYALYLAFLVCAPIFRVBPQSS----APV 416
QY 417 TEKA 420
  417 TEKA 420
Db 421 TANA 424
  421 TANA 424

RESULT 2
DALY_KLEPN STANDARD: PRT: 425 AA.
AC 052718;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-arabinol transporter.
GN DALY.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1033-5p14 / KAY2026;
RX MEDLINE=97464425; PubMed=9324246;
RA Heuel H., Turgut S., Schmid K., Lengeler J.W.;
RT "Substrate recognition domains as revealed by active hybrids between
  the D-arabinol and ribitol transporters from Klebsiella
  pneumoniae."
RL J. Bacteriol. 179:6014-6019(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX
  SUBFAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF045245; AAC26500.1;
DR InterPro: IPR004748; Keloglu permease.
DR Tricore: TIGR00897; 2A0118.1.
KW Transport; Sugar transport; Transmembrane.
FT TRANSMEM 1 7
FT TRANSMEM 8 28 1 (POTENTIAL).
FT TRANSMEM 29 51 1 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 52 72 2 (POTENTIAL).
FT TRANSMEM 73 80 3 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 101 3 (POTENTIAL).
FT TRANSMEM 102 107 4 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 108 128 4 (POTENTIAL).
FT TRANSMEM 129 141 5 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 142 162 5 (POTENTIAL).
FT TRANSMEM 163 172 6 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 173 193 6 (POTENTIAL).
FT TRANSMEM 194 237 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 238 258 7 (POTENTIAL).
FT TRANSMEM 259 263 8 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 264 284 8 (POTENTIAL).
FT TRANSMEM 285 295 8 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 296 316 9 (POTENTIAL).
FT TRANSMEM 317 323 9 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 324 344 10 (POTENTIAL).
FT TRANSMEM 345 360 10 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 361 381 11 (POTENTIAL).
FT TRANSMEM 382 383 11 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 384 404 12 (POTENTIAL).
FT TRANSMEM 405 425 12 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 425 AA: 46826 MW: 5544C2BFA8EC215 CRC64:

Query Match 86.3%: Score 1907; DB 1; Length 425;
Best Local Similarity 84.5%: Pred. No. 6.2e-103;
Matches 354; Conservative 37; Mismatches 28; Indels 0; Gaps 0;

QY 1 MSRRNKQMLGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASFAFTLYGLA 60
  1 MSRRNKQMLGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASFAFTLYGLA 60
Db 1 MSRRNKQMLGLPLHLINGYIAIAVFMGDFELAFLSHYIKSLGFTPAEASFAFTLYGLA 60
QY 61 AALSAMWSGVVAELITTPQKTMILGCVFVHLFVFGQANYGILLFYGIRGLAYP 120
  61 AALSAMWSGVVAELITTPQKTMILGCVFVHLFVFGQANYGILLFYGIRGLAYP 120
Db 61 AALSAMWSGVVAELITTPQKTMILGCVFVHLFVFGQANYGILLFYGIRGLAYP 120
QY 121 LFLYSPFVLIHNVRSNSSAIGWMAVYSGVIGVAGSYIPSTTIPIMGEMGLMALA 180
  121 LFLYSPFVLIHNVRSNSSAIGWMAVYSGVIGVAGSYIPSTTIPIMGEMGLMALA 180
Db 121 LFLYSPFVLIHNVRSNSSAIGWMAVYSGVIGVAGSYIPSTTIPIMGEMGLMALA 180
QY 181 FCFAGVYIAMSIRHVKTPGHHNLTREKFAELSRATVLLTYRNIFLSSIVRIINTLS 240
  181 FCFAGVYIAMSIRHVKTPGHHNLTREKFAELSRATVLLTYRNIFLSSIVRIINTLS 240
Db 181 FCFAGVYIAMSIRHVKTPGHHNLTREKFAELSRATVLLTYRNIFLSSIVRIINTLS 240
QY 241 LFGFVAVTMPMFVDELFTTSEMLOVNAAPFTTIFSNIFGIVAERKGMKRVIRMGCL 300
  241 LFGFVAVTMPMFVDELFTTSEMLOVNAAPFTTIFSNIFGIVAERKGMKRVIRMGCL 300
Db 241 LFGFVAVTMPMFVDELFTTSEMLOVNAAPFTTIFSNIFGIVAERKGMKRVIRMGCL 300
QY 301 GMAASSIAFYMPQYEGHNWMAIPALGTFAAFVPAAPVPALEPKHKAISVYN 360
  301 GMAASSIAFYMPQYEGHNWMAIPALGTFAAFVPAAPVPALEPKHKAISVYN 360
Db 301 GMAASSIAFYMPQYEGHNWMAIPALGTFAAFVPAAPVPALEPKHKAISVYN 360
QY 361 LSGMNSFLAPALAVLLPFFSTIGVYIATYALYLAFLVCAPIFRVBPQSSAPVTEK 419
  361 LSGMNSFLAPALAVLLPFFSTIGVYIATYALYLAFLVCAPIFRVBPQSSAPVTEK 419
Db 361 LSGMNSFLAPALAVLLPFFSTIGVYIATYALYLAFLVCAPIFRVBPQSSAPVTEK 419

```

RESULT 3
CSBX_BACSU STANDARD: PRT: 435 AA.
AC 005390: 032056: 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Alpha-ketoglutarate permease.
GN CSBX.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PY79;
RX MEDLINE=97177783; PubMed=9025289;
RA Gomez M., Cutting S.M.;
RT "bopC encodes a putative forespore regulator of the Bacillus subtilis
RT sigma K checkpoint.";
RL Microbiology 143:157-170(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Tosato V., Bolotin A., Bertani I., Valentino I., Bruschi C.V.;
RT "A 17.8 kb segment in the spoVB-nadC region of the Bacillus subtilis
RT 168 chromosome: sequencing and ruv operon identification.";
RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Barterio M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriotti R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Erlington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
RA Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Setor S.J., Setor P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tashobol V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,
RA Viari A., Mamput R., Medler H., Medler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Darchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP CHARACTERIZATION OF EXPRESSION.
RX MEDLINE=97254445; PubMed=909855;
RA Gomez M., Cutting S.M.;
RT "Identification of a new sigmaB-controlled gene, csbX, in Bacillus
RT subtilis.";
RL Gene 188:29-33(1997).
CC -i- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -i- DEVELOPMENTAL STAGE: EXPRESSED DURING THE STATIONARY PHASE OF CELL
CC GROWTH.
CC -i- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX

CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X93081; CA63620.1; -
CC EMBL: Y15896; CAB75328.1; -
CC EMBL: Z99118; CAB14736.1; -
CC Subtilist: BG11922; csbx.
CC Interpro: IPR004748; ketoglu_permease.
CC Interpro: IPR003662; sub_transporter.
CC Pfam: PF00083; sugar_tr_1.
CC TIGRfam: TIGR00897; 2A0118; 1.
CC Transport; Transmembrane; Complete proteome.
CC CYTOPLASMIC (POTENTIAL).
CC 1 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 2 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 3 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 4 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 5 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 6 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 7 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 8 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 9 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 10 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 11 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 12 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 0 -> E (IN REF. 1).
CC G -> A (IN REF. 1).
CC MISSING (IN REF. 1).
CC F -> S (IN REF. 1).
CC 364
CC SEQUENCE 435 AA; 47863 MW; 408A1A91E019417 CRC64;
SQ
Query Match 37.1%; Score 821; DB-1; Length 435;
Best Local Similarity 36.5%; Pred. No. 1,1e-40;
Matches 146; Conservative 89; Mismatches 161; Indels 4; Gaps 2;
QY 9 LGPLHLIMGYIAIAVFMGTGDFELAFLSHYISLGFTPPAEAFATLYGLAALSAWS 68
DB 14 IGIPSHMWCYIOGVIFMNGDGLGQMLSPFLVDHGSMSQASLFTMGIAVTISAMLS 73
QY 69 GVAVAEITTPKTMIGFVLMCFVHVLFLVGLQOANGILFLFYGRGLAVPLFLYSFY 128
DB 74 GTFVQTMGRPKMTVGLAFITLSAFAIGWAIPIHMYIPALLGSVALRGIDGYPFAVSFLV 133
QY 129 VITHNVRSENSSALGNVNAVYSGVGAGSYIPSPFIPIPMGEMGLTALACFPAGVI 188
DB 134 WVSYSSTSONILGRAVGMFWFMFCGLNVLDLPFYSVAVPAGFINTLMSALLPVAAGIL 193
QY 189 AMISLRHVKTGGMHNTTPREKFAELSBRAVTLLYTNNIFLSSIVRIINTLSFGFAVIM 248
DB 194 ALFFNKDKFTPIOKD---QPKKELSKATTFINENPKVIGGVKTIKINISGFGRAIFL 250
QY 249 PMKVFDELGFTTSEMIQVAAWPAFFTTIFSNIFGVIKAEKGMKRVIRWEGCLGAAMASILA 308
DB 251 P-TYLARVGSVSEMIQIMGTLEFFVNIVFNIIGAVGDKIGMKNVIMWFGVCGGIFTLA 309

QY 309 EYYPQYFGHNYMMAMIPALALGTVAAPVMAFPALBPBKGAISVYNLSAGNSNF 368
DB 310 LLYNPQLIGHQYWMYLMITACCYGAALGYPLSLPTLAPDNKGAAMSVNLGFGICAF 369
QY 369 LAPAIANVLLPWFSTIGVIAATYATLYLAFVLCAFIRVEQ 408
DB 370 IAPGVSLFTGLPGAGGVWIFALYFSAFLTRFETISE 409

RESULT 4
YOAB_BACSU
ID YOAB_BACSU STANDARD: PRT; 414 AA.
AC 034864;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative transporter yoab.
GN YOAB
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN 11
RP SEQUENCE FROM N.A.
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
the tarc and ochab loci cloned in a yeast artificial chromosome";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Aavedo V., Ogassawara N., Moszer I., Albertini A.M., Alloni G.,
RA Borzias R., Boursier L., Brans A., Braun M., Britnell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denton F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Ertlen R.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
RA Gilm S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Potwolik S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlon E., Schleich S., Schroeter R., Scofield F.,
RA Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoli E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viart A., Wambutt R., Wedler E., Wedler H., Welteregger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis".
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. CSBX
CC SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC EMBL: AF027868; AAB84444.1;
DR EMBL: 299114; CAB13747.1;
DR SUBLIST: BG13473; yoab.
DR InterPro: IPR004748; Keloglu-permease.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1;
DR TIGRFAMs: TIGR00897; 2A0118; 1;
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT DOMAIN 1 11
FT TRANSSEM 12 32
FT DOMAIN 33 47
FT TRANSSEM 48 68
FT DOMAIN 69 75
FT TRANSSEM 76 96
FT DOMAIN 97 107
FT TRANSSEM 108 128
FT DOMAIN 129 136
FT TRANSSEM 137 157
FT DOMAIN 158 167
FT TRANSSEM 168 188
FT DOMAIN 189 219
FT TRANSSEM 220 240
FT DOMAIN 241 255
FT TRANSSEM 256 276
FT DOMAIN 277 286
FT TRANSSEM 287 307
FT DOMAIN 308 316
FT TRANSSEM 317 337
FT DOMAIN 338 343
FT TRANSSEM 344 364
FT DOMAIN 365 375
FT TRANSSEM 376 398
FT DOMAIN 399 414
SQ SEQUENCE 414 AA: 45376 MW: 4FAELICBCBDD744C4 CRG64;
Query Match 32.18; Score 709.5; DB 1; Length 414;
Best Local Similarity 33.6%; Pred. No. 2.5e-34;
Matches 139; Conservative 88; Mismatches 172; Indels 15; Gaps 4;
QY 9 LGDPLHIMGYIAIVMTGDFLAFSLHYISIGTPAPAFPLVGLAALSAWS 68
DB 5 IGIPKRLAKGFLVGLVLEPMADGLEQGLSPFLIENGLTVQOQSIRSYIALAIAWS 64
QY 69 GYVAEITTPKMTLIGFVLVCVPHVLEVLGQANGLILEFYGRLAVPLFLYSFIV 128
DB 65 GVCLEAFGARKTMEGGLFVYGTAFIVGFEQNLPPVYVYVYVGLGYPFAVSFLT 124
QY 129 VTIHNVSESSSALGYNVAVSYGIVASYSIPSPFIPIMGEGTILMALACFAGVI 188
DB 125 WVIYRTPQSKLSTAVGFWFVAYCIGMFVGAWSVYAKAFGLNLTMSIFWVCIAFF 184
QY 189 AMISLRHVTKPRGHNNHNTPREKFA-ELSRVATLLYTRNFLSIVRIINTLSLGFVAVI 247
DB 185 ALF-----INNDREKKRRKRSERAEELKGVITLFTPRVLGTGIRIINSIGTGPV 240
QY 248 MPKWFDELGFTTSEWLVMAAFPTTIFSNIFMGVIAERKGMWRVIRWEGCLGMAASSL 307
DB 241 LPMHMAQH-GISTNVMLDINGTIFLGNIVNLFGLVIGDKFGKNTVIMFGVGGCIFTV 299
QY 308 AFYYPQYFGHNYMMAMIPALALGTVAAPVMAFPALBPBKGAISVYNLSAGNSNF 367
DB 300 LLYNPQLIGHQYWMYLMITACCYGAALGYPLSLPTLAPDNKGAAMSVNLGFGICAF 359
QY 368 LAPAIANVLLPWFSTIGVIAATYATLYLAFVLCAFIRVEQ 421
DB 360 FVGPALAMFLTGLVGAQGVWIFALYFSAFLTRFETISE-----EKAV 404

RESULT 5
UHPT_CHLTR
ID UHPT_CHLTR STANDARD: PRT; 456 AA.

```

AC 084548;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hexose phosphate transport protein.
GN CT344.
OS Chlamydia trachomatis.
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
ON NCBI_TaxID=813;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Katman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.,
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE001325; AAC68146.1;
DR InterPro: IPR000849; GLPT_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRPFAM: TIGR00881; 2A0104; 1.
DR PROSITE: PS00942; GLPT; 1.
KW Transport; Sugar transport; Transmembrane; Complete proteome.
FT TRANSMEM 34
FT TRANSMEM 54
FT TRANSMEM 70
FT TRANSMEM 90
FT TRANSMEM 113
FT TRANSMEM 133
FT TRANSMEM 161
FT TRANSMEM 181
FT TRANSMEM 185
FT TRANSMEM 205
FT TRANSMEM 257
FT TRANSMEM 277
FT TRANSMEM 302
FT TRANSMEM 322
FT TRANSMEM 331
FT TRANSMEM 351
FT TRANSMEM 363
FT TRANSMEM 383
FT TRANSMEM 394
FT TRANSMEM 414
FT TRANSMEM 421
FT TRANSMEM 441
SQ SEQUENCE 456 AA; 51729 MW; E443610D6B5B7F8F CRC64;
Query Match 8.2%; Score 180.5; DB 1; Length 456;
Best Local Similarity 22.3%; Pred. No. 0.00056;
Matches 105; Conservative 71; Mismatches 171; Indels 123; Gaps 25;
OY 1 MSNNNNQW-LGLPLHLIMGIAIAVEMTGDFELAFSLHIKSLGTFPAPAS-FATFLYG 58
DB 23 VKQYTYWRRIRIYVSMFLGVFF--YFTRKSTFPA-MPTLIADLGDKADLGIIIGSTLY- 78
OY 59 LAAALSAWVGVAEITTPKTMKLGVCVFLVFLVGLGOANYGLILFYGI----- 114
DB 79 ITYGISKFSVGSVDOSNPRYFMAIGLIIGISNIF---FGL-SRTPLFLVLEWGINGF 134
OY 115 RGLAYP---LFLYSFIVIIHNVRSENSSALGWYAVYSGVIGVASYIPSTF---IP 167
DB 135 QCGMGPCARALLTHWY-----SKSERGTWMSVSTSHNIGCALIPVLTVGAID 182
OY 168 IMKEMGTMLALAFCAAGVIAIMISLRHVTPGHMNTPREFAELSRVATLLYTRNT 227
DB 183 YTGWRAMPAPGIICTIIMGFTLIDRLR--DTPQSL-GLPAIEKFRKEEDAHPEETTD 239
OY 228 FLSSIVRIINT-----LSLGFVAIVPMFVDELGFTTSEMQLQWMAAF 271

```

```

DB 240 LEEBAERELSTKEILFTYVLSNKMWLFSPASFITYVRAVND-----MSALY 288
OY 272 -----FTTIFSNI-----FWGIVAE-----KAGMRYIRFGLGMAAS 306
DB 289 LIETKDYSTVKNALCVSLFEIGLFGMLAGWLSDTISKGRGPMNVY---FSLGLVSI 345
OY 307 LAEYVMPDYGHNVMMAMPAIALGFFVAFVMAAV-FPALPKRKGALISYVNSAGM 365
DB 346 LGLMGTRDYF--VWMDGTFLFIIGFLL--FGQMMIGLAAELSHKAA-----GRASGF 397
OY 366 SNEFLAPAAVALLPWFSTIGVIAAYTAL-----YLAEFVLCAT 404
DB 398 TG-----WFAVFCAPAGYPLGRVAQDWGMHGFEVALLLACALI 435
RESULT 6
NORA_STAM
ID NORA_STAM STANDARD; PRT; 388 AA.
AC P21191;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Quinolone resistance protein nora.
GN NORA OR SAV0695 OR SA0650 OR MM0657.
OS Staphylococcus aureus (strain MU50 / ATCC 700699),
OS Staphylococcus aureus (strain N315),
OS Staphylococcus aureus (strain MM2), and
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
ON NCBI_TaxID=158878, 158879, 196620, 1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MU50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kaneshita M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT aureus."
RL Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MM2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Ikama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.,
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA."
RL Lancet 359:1819-1827(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91072245; PubMed=2174864;
RA Yoshida H., Bogaki M., Nakamura S., Ubukata K., Konno M.,
RT "Nucleotide sequence and characterization of the Staphylococcus
RT aureus nora gene, which confers resistance to quinolones."
RL J. Bacteriol. 172:6942-6949(1990).
RN [4]
RP SEQUENCE OF 230-388 FROM N.A.
RC STRAIN=209P;
RX MEDLINE=91058531; PubMed=2173911;
RA Oshita Y., Hiramatsu K., Yokota T.,
RT "A point mutation in nora gene is responsible for quinolone
RT resistance in Staphylococcus aureus."
RL Biochem. Biophys. Res. Commun. 172:1028-1034(1990).
CC -1- FUNCTION: INVOLVED IN QUINOLONE RESISTANCE. MAY CONSTITUTE A
CC MEMBRANE-ASSOCIATED ACTIVE EFFLUX PUMP OF HYDROPHILIC QUINOLONES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

```

```

CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003360; BAB56857.1; -
DR EMBL: AP003131; BAB41883.1; -
DR EMBL: AP004824; BAB94522.1; -
DR EMBL: D90119; BAA14147.1; -
DR EMBL: M62960; AAB59089.1; -
DR PIR: A37838; A37838.
DR InterPro: IPR004734; Drug_resist.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRfams: TIGR00880; 2_A_01_02; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 353 375 POTENTIAL.
FT TRANSMEM 352 352 D -> A (NO QUINOLONE RESISTANCE).
FT VARIANT 352 352 M -> V (IN REF. 4).
FT CONFLICT 362 362 A -> D (IN REF. 4).
FT SEQUENCE 388 AA; 42265 MW; C36177EAAADAC57 CRC64;

Query Match 7.9%; Score 174; DB 1; Length 388;
Best Local Similarity 20.5%; Pred. No. 0.0012;
Matches 85; Conservative 76; Mismatches 191; Indels 62; Gaps 15;

```

```

UHPT_CHLPN STANDARD; PRT; 455 AA.
ID UHPT_CHLPN
AC Q927N9; Q9J019;
DR 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hexose phosphate transport protein.
GN CPN0665 OR CPN082.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
ON NCBI_Taxid=83558;
RX MEDLINE-99206606; PubMed-10192388;
RA Kaiman S., Mitchell W., Matathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RA Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001649; AAD18804.1; -
DR EMBL: AE002171; AAF37968.1; -
DR EMBL: AP002547; BAA98872.1; -
DR TIGR: CP0082;
DR InterPro: IPR000849; G1PT_transporter.
DR TIGRfams: TIGR00881; 2A0104; 1.
DR PROSITE: PS00942; GLPT; 1.
KW Transport; Sugar transport; Transmembrane; Complete proteome.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT SEQUENCE 455 AA; 51659 MW; F6404E2F1F4CA076 CRC64;

```

RESULT 7


```

OS Acinetobacter calcoaceticus.
OC Bacteri; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BD413 / ADPI;
RX MEDLINE=97440147; PubMed=9294455;
RA Williams P.A., Shaw L.E.;
RT "muck, a gene in Acinetobacter calcoaceticus ADPI (BD413), encodes the
RT ability to grow on exogenous cis,cis-muconate as the sole carbon
RT source";
RL J. Bacteriol. 179:5935-5942(1997).
CC -1- FUNCTION: PROBABLE UPTAKE OF MUCONATE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 087258; AAC27117.1;
DR InterPro: IPR004746; Bnzte_transport.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR TIGRfam: TIGR00885; 2A0115.1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 1 (POTENTIAL).
FT DOMAIN 38 53 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 54 74 2 (POTENTIAL).
FT DOMAIN 75 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 106 3 (POTENTIAL).
FT DOMAIN 107 112 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 113 133 4 (POTENTIAL).
FT DOMAIN 134 145 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 146 166 5 (POTENTIAL).
FT DOMAIN 167 171 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 172 192 6 (POTENTIAL).
FT DOMAIN 193 228 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 229 249 7 (POTENTIAL).
FT DOMAIN 250 266 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 267 287 8 (POTENTIAL).
FT DOMAIN 288 293 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 294 314 9 (POTENTIAL).
FT DOMAIN 315 318 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 319 339 10 (POTENTIAL).
FT DOMAIN 340 352 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 353 376 11 (POTENTIAL).
FT DOMAIN 377 382 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 383 403 12 (POTENTIAL).
FT DOMAIN 404 413 CYTOPLASMIC (POTENTIAL).
SO SEQUENCE 413 AA: 45245 MW: 2D88CE31C45CC65 CRC64;
Query Match 7.1%; Score 158; DB 1; Length 413;
Best Local Similarity 24.6%; Pred. No. 0.01;
Matches 110; Conservative 63; Mismatches 192; Indels 82; Gaps 25;

```

```

OY 104 --NGXLLLEFGYGRGLAYPLFLY-SFTVYIIHNVRENSSSALGWVAVYSGIGVAGSY 160
OY 105 FIOFG-VLRFEPASIGLG---SLYACNTLMAEVPYRTYRTVLGTLOAGTWGVYA-TL 159
OY 161 IPSTFTIPNEMGTMLALPCFAGSVIAMI SLRHVTP-----GHMINTLPREKFEEL-S 215
OY 160 LAGWLIPDHG-----WRLV-FYVAIIPVMAVHFFVPPEPAAQOSRLAPSKOTETVKT 213
OY 216 RAVVLLY---TNRNIF-TSSIVRIINTLSFGFAVIRPMHFVDELGFTSEWLVMAAF 271
OY 214 SAFELIQRDKNRRNMFLMTALTAGFLQFGYYGVNMMPSVLSLQKFKEMRAYMVGTY 273
OY 272 FTTIFSNIFNGIYAEKGMKRVIRMGCLMA--ASSLAFY-----YMPQFGRHYNM 322
OY 274 TAMILGKILGAFMADKLG-RRETYAFGAICTAIFLPIVEYNSPDNIIYLVLVEFLYG- 331
OY 323 AMIPALIGFVAFAFVMAVFPALPEKHGAISVYNLSAGMSNFLAPAVVLLPWFPS 382
OY 332 --IPYGVNATVMTESFPTAIRGTAI-----GGAIVNGRLGAA---IAPTIGFLASGGS 380
OY 383 T-IGVVIATYALYLLAFVLCATFIRVEQ 408
OY 381 IGLGFVVMGAAYFICGVIPALFIKEKQ 407
Db
RESULT 10
YX10_BACSU
ID YX10_BACSU STANDARD; PRT; 428 AA.
AC P42306;
DT 01-NOV-1995 (Rel. 32, Last Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yx10.1
GN YX10 OR S3AR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / BGSJAL;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the ltc and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacxy region.";
RL Microbiology 142:3113-3123(1996).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Bruns A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel N.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Eriington J., Fabret C., Ferrari E., Fougere D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shln B.S., Soldo B.,

```


RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsita P., Tognoni A.,
RA Tosto V., Uchiyama S., Vandenbol M., Vannier P., Vassaretto A.,
RA Viati A., Wamutit R., Wedler E., Wedler H., Wetzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RT Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: SOME, TO YEAST YCD8 AND S.POMBE SPAC2G11.13.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: D83026; BAA11694.1; -
DR EMBL: 299124; CAB15946.1; -
DR Subtilisin; Bg11144; yx10.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 238 258 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 392 412 POTENTIAL.
SQ SEQUENCE 428 AA; 47272 MW; B92AB9C6B544C01 CRC64;

Query Match 6.9%; Score 151.5; DB 1; Length 428;
Best Local Similarity 22.0%; Pred. No. 0.024;
Matches 98; Conservative 82; Mismatches 184; Indels 81; Gaps 22;

RESULT 11
TCR_STRAG STANDARD; PRT; 458 AA.
AC P13924;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Tetracycline resistance protein.
GN TET.
OS Streptococcus agalactiae.
OC Plasmid pMW158.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus;
OX NCBI_TaxID=1311;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-90016790; PubMed-2677995;
RA van der Lelje D., Bron S., Venema G., Oskam L.;
RT "Similarity of minus origins of replication and flanking open reading
RT frames of plasmids PUB110, PTB913 and pMW158.";
RL Nucleic Acids Res. 17:7283-7294(1989).
CC -1- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE
CC ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTI-PORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X15669; CAA33712.1; -
DR PIR: C25599; YTSOC;
DR InterPro: IPR001411; TCR_TetB.
DR PRINTS: PR01036; TCR_TetB.
KW Antibiotic resistance; Transmembrane; Transport; Symport; Plasmid.
FT TRANSMEM 12 33 POTENTIAL.
FT TRANSMEM 81 100 POTENTIAL.
FT TRANSMEM 111 129 POTENTIAL.
FT TRANSMEM 140 162 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 223 240 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 346 365 POTENTIAL.
FT TRANSMEM 432 451 POTENTIAL.
SQ SEQUENCE 458 AA; 50006 MW; AD2014E7CA19995 CRC64;

Query Match 6.5%; Score 144; DB 1; Length 458;
Best Local Similarity 19.7%; Pred. No. 0.069;
Matches 84; Conservative 83; Mismatches 144; Indels 116; Gaps 19;

QY 181 FCFAGVIAISLRHV-----TPGHMNLPRKFAELSAVTLTYTNRIFFLSIV 233
 Db 225 FLIYVSLEFLFVYKHRIKRDPEVDPGIGKNI-----LFMIGVL 263
 QY 234 ---RIIMTSLFSGFAVIMPMFVDELGFTTSEMLOVMAAFPTTIFSNIF---GIVAE 286
 Db 264 CGGIIFCTVA--GFVSNVPPYMKDVHQLSTAE---IGSVIIFPGTMSVITFGYIGILVD 318
 QY 287 KMGMYIRMEGCGIAGMAASSLAFFYMPQFCHNYMAMIPAIAG--TFVAAPVMAAFV 344
 Db 319 RRGVLVLN-IGVTFLSVSLTAFLE--TTSWFMITIIYVGGLSFTKTVIP-TVVS 374
 QY 345 PALEPKHGAIVSYNLSAGSNFLAPAAVY-----LDPWFSTIGV---VIA 389
 Db 375 SSLKQDAGAGMSLINFSTFSE--GTGIAIVGGLSLPILDQRLP---MGVDQSTYL 428
 QY 390 YTAIXYL 396
 Db 429 YSNLLL 435

RESULT 12

BETP_CORGL STANDARD; PRT; 595 AA.
 AC P54582;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glycine betaine transporter betP.
 GN BETP OR CGL0892.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.
 NC NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Peter H. J.
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S. J.
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: HIGH-AFFINITY UPTAKE OF GLYCINE BETAINES (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE BCCP (TC 2.33) FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X93514; CAA6371.1; -
 DR EMBL: AP005276; BAB98285.1; -
 DR InterPro: IPR000060; BCCP_transporter.
 DR Pfam: PF02028; BCCP_1.
 DR ProDom: PD010111; BCCP_transporter; 1.
 DR TIGRFAMs: TIGR00842; bccp1.
 DR PROSITE: PS01303; BCCP_1.
 KM Transport: Transmembrane.
 FT TRANSMEM 60 80
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 139 159 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.

FT TRANSMEM 310 330 POTENTIAL.
 FT TRANSMEM 366 386 POTENTIAL.
 FT TRANSMEM 397 417 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 FT TRANSMEM 490 510 POTENTIAL.
 FT TRANSMEM 521 541 POTENTIAL.
 SQ SEQUENCE 595 AA; 64209 MW; 8354B41C834CEEE2 CRC64;

Query Match 6.58; Score 143.5; DB 1; Length 595;
 Best Local Similarity 20.54; Pred. No. 0.091;
 Matches 105; Conservative 70; Mismatches 154; Indels 183; Gaps 26;

QY 20 IAIIVMTGDFELAPLSHYKSLGFTPAEASFAFTLYGLAALSAWVSGVAEITTPK 79
 Db 69 IVALTVWVGIGCFKDSFTNFSSALSAVDNLGMAFILFG---TVVFVFVYIAASKFG 123
 QY 80 TMLG-----FVLMCPHVFLVFGIGQANVYGLLFLYGRIGLAYPLFLSYLV- 129
 Db 124 TIRLGIDEAPEFTVSW---ISMFAAGM--GIGLMFY--GTPEPLTFYRNGVPG 172
 QY 130 -IHNVRSESSSALGW---WAVYSVIGVAGSY-----IPSTIPIMGKGTL 175
 Db 173 HDEHNVGAMSTYMFHMTLHPMAIYAL-VGLAIYSTFRVGRKQLSSAVPLIGEGAE 231
 QY 176 -WL-----ALAF--CFAG-----GVIAISLRHV 196
 Db 232 GWLGKIDILAIATVFGTCSLGLGALQAGLSAANTIEDPSDWTIVGIVSVTLAFI 291
 QY 197 KTPGHMNLPRKFAELS---RAVTLLTYNRNIFLSYRII----- 236
 Db 292 ---FSAISGVKGIQYL-SNANNVLAALLAIFEVVGPVYSIINLLPGS 336
 QY 237 --NTLSFGFAVIMPMFVDELGFTTSEMLOVMAAFPTTIFS-NIPMGIVAEKMGMY 293
 Db 337 IGNLYSNFPQAGRTAMSADG---TAGEWLGSWTFIYMAWIMSPVGNFLAIRISGRS 393
 QY 294 IR-----WFGCLGMAASSLAFFYMPQFCHNYV----- 321
 Db 394 IREFILGLLVPAQSVTVSIFG--GTATVF---FQNGESIWGDGAEEQLGLLHAL 447
 QY 322 ---MAMITPAALGCTVAAFVPMAAVFPALPEPKH-----KGAISVYNL 361
 Db 448 PGQIMGIIITAMILLGTFTTTSADSASTVGMGTSGHGLEANKVYTAAMGATATIGTLTL 507
 QY 362 SAGMSNFLAPAAVAVLL---PW-FSTIGVYIA 389
 Db 508 LSGGDMLNSLNQNTYIVAAAPFLFVIGLWFA 539

RESULT 13
 YAJR_ECOLI STANDARD; PRT; 454 AA.
 ID YAJR_ECOLI
 AC P77726;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical transport protein yajR.
 GN YAJR OR B0427.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y. J.
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Sequence 277:1453-1474(1997);
 RN [2]

[illegible][illegible]

```
FT TRANSNM 225 243 PROBABLE.
FT DOMAIN 244 254 EXTRACELLULAR (PROBABLE).
FT TRANSNM 255 273 PROBABLE.
FT DOMAIN 274 293 CYTOPLASMIC (PROBABLE).
FT TRANSNM 294 312 PROBABLE.
FT DOMAIN 313 335 EXTRACELLULAR (PROBABLE).
FT TRANSNM 336 354 PROBABLE.
FT DOMAIN 355 372 CYTOPLASMIC (PROBABLE).
FT TRANSNM 373 391 PROBABLE.
FT DOMAIN 392 425 EXTRACELLULAR (PROBABLE).
FT TRANSNM 426 444 PROBABLE.
FT DOMAIN 445 497 CYTOPLASMIC (PROBABLE).
FT CONFLICT 105 105 I -> V (IN REF. 1).
FT CONFLICT 219 219 S -> A (IN REF. 1).
FT CONFLICT 222 222 R -> K (IN REF. 1).
FT CONFLICT 226 226 L -> I (IN REF. 1).
FT CONFLICT 233 233 V -> L (IN REF. 1).
FT CONFLICT 237 237 L -> I (IN REF. 1).
FT CONFLICT 243 243 L -> I (IN REF. 1).
FT CONFLICT 323 323 D -> N (IN REF. 1).
FT CONFLICT 329 329 I -> F (IN REF. 1).
FT CONFLICT 356 356 L -> I (IN REF. 1).
FT CONFLICT 362 362 E -> D (IN REF. 1).
FT CONFLICT 380 380 I -> A (IN REF. 1).
FT CONFLICT 458 458 L -> I (IN REF. 1).
FT CONFLICT 485 485 V -> I (IN REF. 1).
SQ SEQUENCE 497 AA; 54844 MW; 348AE0D27960610C CRC64;
```

Query Match 6.4%; Score 141; DB 1; Length 497;
Best Local Similarity 19.2%; Pred. No. 0.11;

Matches 93; Conservativity 79; Mismatches 163; Indels 150; Gaps 21;

```
QY 1 MSRRNNKWLGLPLHLI-----WGYIAVPMGDFELAFLSHYKSLGFTP 47
DB 4 LNKTEKTFEFOPRDLTLFOTEFEMERFSYGMRAILVY-----YLVALTTADNAGLPLK 58
QY 48 AEASFATLGLAALASWYGVAAE-IITPOKMLIGFVLMCFVH-LFVLEGLQANY 105
DB 59 AQMAATYITGALVYSTIYGWVADRGLGASRTIFLGILITLGHIALTPFELSS--- 115
QY 106 GLILFYGIRGLAVPLFLYSFIVIIHNVSSENSSALGWYMA-----VYSVGI 154
DB 116 -----LFVALFLIITIGTGMKLPNTSNMVGHLXSDDSRRDTGFIIVVGI 160
QY 155 GVAGSYTPTPIPMGEMGLTMLALFCFAGGVITAMTS-----LRHY-----KTGCHMHN 205
DB 161 NM-GSLIAPLIYGVGGVNVHLGFSLSAIGMIFALFAYWYGRHRHPEIGREPSNPMDS 219
QY 206 TPREFAEISRAVTLTYTNRIPL-----SIVRIITLSLGFVIMPMFVDELGF 258
DB 220 KARNFPL-IITLIYVIAIIGFELLYQASPAFNFNNFINVLSTIG--IVPFIIF-VMMF 275
QY 259 TTSE-----WLOVMAFFETTTISNIFMGIVAEKMGKMRV-----IRMFGLGM 302
DB 276 TSKVESDERRRKLTAVIPLFLSAI---VFMAIEQSSTIIAVWGESRSNLDPTWFG--- 328
QY 303 AASLAFYVMPYF-----GHIYWM 322
DB 329 ----ITFHIDPSWQOLLNPLFIVLLSPITFVRLMNLGGEROSTIVKGLGLMLTGISYLI 384
QY 323 AMIPALIGTEVAA-----FVPMMAVFPALBPKHGAISIVYNLSAGSNFLAPAI--AVV 376
DB 385 MTLPGLLNGISGRASALMLVLMFAVOMAGE-----LLVSPVGLSVSTKLAPAFOSOM 437
QY 377 LLEWF 381
DB 438 MAMWF 442
```

RESULT 15
G6PU_HUMAN STANDARD; PRT; 429 AA.
ID G6PU_HUMAN
AC Q43826; Q9U119; Q9UN54; Q96016;

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose 6-phosphate translocase (Glucose 5-phosphate transporter)
DE (PRO0685)
GN G6PT1 OR G6PT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT GSD-IB CYS-339.
RC TISSUE=Urinary bladder;
RX MEDLINE=9808917; PubMed=9428641;
RA Gerin I., Veiga-Da-Cunha M., Achouri Y., Collet J.-F.,
RA van Schaftingen E.;
RT "Sequence of a putative glucose 6-phosphate translocase, mutated in
RT glycogen storage disease type Ib.";
RL FEBS Lett. 419:235-238(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT GSD-IB ARG-118.
RX MEDLINE=99072316; PubMed=9856496;
RA Ihara K., Kuromaru R., Hara T.;
RT "Genomic structure of the human glucose 6-phosphate translocase gene
RT and novel mutations in the gene of a Japanese patient with glycogen
RT storage disease type Ib.";
RL Hum. Genet. 103:493-496(1998).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
RX MEDLINE=99150335; PubMed=10026167;
RA Hiratawa H., Pan C.-J., Lin B., Moses S.W., Chou J.Y.;
RT "Inactivation of the glucose 6-phosphate transporter causes glycogen
RT storage disease type Ib.";
RL J. Biol. Chem. 274:5532-5536(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal liver;
RA Zhang C., Yu Y., Zhang S., Ouyang S., Luo L., Wei H., Zhou G.,
RA Zhang Y., Liu M., He F.;
RT "Functional prediction of the coding sequences of 9 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Li Y., van de Werve G.;
RT "Four different transcripts of putative glucose-6-phosphate
RT translocase in human leukocytes.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99148010; PubMed=10023055;
RA Gerin I., Veiga-Da-Cunha M., Noel G., Van Schaftingen E.;
RT "Structure of the gene mutated in glycogen storage disease type Ib.";
RL Gene 227:189-195(1999).
RN [7]
RP SEQUENCE FROM N.A.
RX PubMed=10323254;
RA Janecek A.R., Bosshard N.U., Mayatepek E., Schulze A., Gitzelmann R.,
RA Burchell A., Bartem C.R., Janssen B.;
RT "Molecular diagnosis of type Ic glycogen storage disease.";
RL Hum. Genet. 104:275-277(1999).
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Colon, and Lung;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [9]
RP VARIANT GSD-IB ARG-118.
RX MEDLINE=98342107; PubMed=9675154;
RA Kure S., Suzuki Y., Matsubara Y., Sakamoto O., Shintaku H.,
RA Ieshiki G., Hoshida C., Izumi I., Sakura N., Narisawa K.;
RT "Molecular analysis of glycogen storage disease type Ib:"

RT identification of a prevalent mutation among Japanese patients and
RT assignment of a putative glucose-6-phosphate translocase gene to
RT chromosome 11." ;
RL Biochem. Biophys. Res. Commun. 248:426-431(1998).
RN [10]
RN VARIANTS GSD-IB ARG-118 AND VAL-235 DEL.
RP PubMed-10482875;
RA Hou D.-C., Kure S., Suzuki Y., Hasegawa Y., Hara Y., Inoue T.,
RA Kida Y., Matsubara Y., Narisawa K. ;
RT "Glycogen storage disease type Ib: structural and mutational analysis
RT of the microsomal glucose-6-phosphate transporter gene." ;
RL Am. J. Med. Genet. 86:253-257(1999).
RN [11]
RP VARIANT GSD-IB GLU-149.
RA Lam C.-W., Tong S.-F., Lam Y.-Y., Chan B.-Y., Ma C.-H., Lim P.-L. ;
RT "Identification of a novel missense mutation (G149E) in the glucose-6-
RT phosphate translocase gene in a Chinese family with glycogen storage
RT disease Ib." ;
RL Hum. Mutat. 13:507-507(1999).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum (potential).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2;
CC are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: MOSTLY EXPRESSED IN LIVER AND KIDNEY.
CC -1- DISEASE: DEFECTS IN G6PT1 ARE THE CAUSE OF GLYCOGEN STORAGE
CC DISEASE IB (GSD-IB).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 128.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y15409; CAA75608.1; -.
DR EMBL: AF078163; AAC72916.1; -.
DR EMBL: AF097813; AAD19898.1; -.
DR EMBL: AF111852; AAF16691.1; ALT_FRAME.
DR EMBL: AF110819; AAF37735.1; -.
DR EMBL: AF110820; AAF37736.1; -.
DR EMBL: Y17864; CAA7698.1; -.
DR EMBL: AF116864; AAD13111.1; -.
DR EMBL: AF116862; AAD13111.1; JOINED.
DR EMBL: AF116863; AAD13111.1; JOINED.
DR EMBL: BC002400; AAH02400.1; -.
DR EMBL: BC003589; AAH03589.1; -.
DR EMBL: BC014663; AAH14663.1; -.
DR GeneW: HGNC:4061; G6PT1.
DR MIM: 602671; -.
DR MIM: 232220; -.
DR InterPro: IPR000849; G1PT_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR TIGRFAMs: TIGR00881; taO104.1.
DR PROSITE: PS00942; GLPT.1
KW Transmembrane; Transport; Sugar transport; Endoplasmic reticulum;
KW Alternative splicing; Glycogen storage disease; Disease mutation.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 167 187 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 368 388 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT VASAPLIC 328 328 K->KVAFPTLALPLAELTGTEHE (IN ISOFORM
2).

[illegible]

```
Search completed: March 13, 2003, 16:53:47
Job time : 11.4238 secs
```

2)

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 16:43:37 ; Search time 27.0265 Seconds

(without alignments)
3232.531 Million cell updates/sec

Title: US-09-802-208b-5

Perfect score: 2211

Sequence: 1 MSRRNKQWGLPLHLINGVI.....RVEQPFSSAPVTEKALNIS 424

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriap:*
- 17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2211	100.0	424	2 O9F4L5	O9F4L5 escherichia
2	625	28.3	408	16 O8ZK49	O8ZK49 salmonella
3	181	8.2	462	16 O9HWG1	O9HWG1 pseudomonas
4	176	8.0	456	16 O8ZC53	O8ZC53 yersinia pe
5	175	7.9	388	2 O53459	O53459 staphylococ
6	174	7.9	388	2 O03325	O03325 staphylococ
7	173.5	7.8	444	16 O34502	O34502 bacillus su
8	168.5	7.6	457	5 O96186	O96186 plasmidium
9	164.5	7.4	386	17 O58955	O58955 methanococ
10	162.5	7.3	426	16 O8ZQ32	O8ZQ32 salmonella
11	162.5	7.3	426	16 O8Z7M8	O8Z7M8 salmonella
12	161.5	7.3	400	16 P94376	P94376 bacillus su
13	161.5	7.3	411	16 O8ZPR3	O8ZPR3 salmonella
14	159.5	7.2	428	16 O8ZPG6	O8ZPG6 salmonella
15	159.5	7.2	428	16 O8Z716	O8Z716 salmonella
16	158.5	7.2	387	16 O9CJ77	O9CJ77 lactococcus

17	155.5	7.0	419	16 O9HWT2	O9HWT2 pseudomonas
18	155	7.0	397	16 O9A129	O9A129 streptococ
19	155	7.0	412	17 O9HXR5	O9HXR5 halobacteri
20	154.5	7.0	380	17 P95885	P95885 sulfolobus
21	154	7.0	388	2 O9ZNA9	O9ZNA9 staphylococ
22	153	6.9	405	2 O66181	O66181 arthrobacte
23	153	6.9	405	16 O8Z257	O8Z257 salmonella
24	152.5	6.9	392	2 O9L804	O9L804 pseudomonas
25	152	6.9	396	16 O06473	O06473 bacillus su
26	151	6.8	413	16 O9K788	O9K788 bacillus su
27	151	6.8	419	16 O8ZLE4	O8ZLE4 salmonella
28	150.5	6.8	409	16 O9PTL7	O9PTL7 deinococcus
29	150	6.8	401	16 O9ZCF5	O9ZCF5 listeria in
30	150	6.8	437	16 O9HXP5	O9HXP5 pseudomonas
31	149.5	6.8	425	16 O99RC3	O99RC3 staphylococ
32	149.5	6.8	567	16 O9RTS9	O9RTS9 deinococcus
33	148.5	6.7	402	16 O912B6	O912B6 pseudomonas
34	148	6.7	432	10 O9FKV1	O9FKV1 arabidopsis
35	148	6.7	905	2 O93FW1	O93FW1 paracoccus
36	147	6.6	391	17 O979S3	O979S3 thermoplas
37	147	6.6	401	16 O8Y7M4	O8Y7M4 listeria mo
38	147	6.6	431	2 O9ZRF4	O9ZRF4 campylobact
39	147	6.6	431	16 O9PMS5	O9PMS5 campylobact
40	146.5	6.6	436	16 O9XOF0	O9XOF0 thermotoga
41	146.5	6.6	524	16 O98GA1	O98GA1 rhizobium 1
42	146.5	6.6	577	16 O9L134	O9L134 streptomyce
43	146	6.6	387	17 O8U4E5	O8U4E5 pyrococcus
44	145.5	6.6	381	17 O58638	O58638 pyrococcus
45	145	6.6	391	2 O8VU72	O8VU72 lactococcus

ALIGNMENTS

RESULT 1

O9F4L5 PRELIMINARY; PRT; 424 AA.

AC O9F4L5; 01-MAR-2001 (TREMREL. 16, Created)

DT 01-MAR-2001 (TREMREL. 16, Last sequence update)

DE 01-JUN-2002 (TREMREL. 21, Last annotation update)

DE Rbltol transporter.

GN RPTT.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI_TaxId=562;

RP SEQUENCE FROM N.A.

RC SPRAIN-C;

RA LaFayette P.R.; Parrott W.A.;

RT "A non-antibiotic marker for amplification of plant transformation vectors in E. coli."

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY005817; AAG01885.1; -

DR InterPro: IPR001064; Crystallin.

DR InterPro: IPR004748; Keloglu-permease.

DR TRGPRMS; TRGPR0897; 2A0118.1

DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.

DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.

SQ SEQUENCE 424 AA; 46877 MW; BC/FEFPC/7A83962 CRC64;

Query Match 100.0%; Score 2211; DB 2; Length 424;

Best Local Similarity 100.0%; Pred. No. 4.4e-150;

Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSRRNKQWGLPLHLINGVIATAVFMGTGDFELAFSLHYKSLGFTPAEASFAFTLYGLA 60

DB 1 MSRRNKQWGLPLHLINGVIATAVFMGTGDFELAFSLHYKSLGFTPAEASFAFTLYGLA 60

OY 61 AALSAMVSGVAELITPQKTMILGFLVIMCYFHVILFLVFGIGQANYGILLFYGRGLAYP 120

DB 61 AALSAMVSGVAELITPQKTMILGFLVIMCYFHVILFLVFGIGQANYGILLFYGRGLAYP 120

QY 121 LFLYSEIVYIIHNVRSESSSALCWYAVSVGIGVAGSYIPSTPIPMGEMGLMLALA 180
DB 121 LFLYSEIVYIIHNVRSESSSALCWYAVSVGIGVAGSYIPSTPIPMGEMGLMLALA 180
QY 181 FCFAGVYIAMIISLRHVTPGHNHLPREKPAELSRATVTLTYNRIIFLSSIVNIITLS 240
DB 181 FCFAGVYIAMIISLRHVTPGHNHLPREKPAELSRATVTLTYNRIIFLSSIVNIITLS 240
QY 241 LFGFAVIMPMVDELGFTTSEMLQVMAAFETTFISNIFMGIVAEKKGMVRIRWEGCL 300
DB 241 LFGFAVIMPMVDELGFTTSEMLQVMAAFETTFISNIFMGIVAEKKGMVRIRWEGCL 300
QY 301 GMASSIAFYMPQYFGHNHNMAMIPALIGTFYAAAFVPMMAVPALEPKHKAISVYN 360
DB 301 GMASSIAFYMPQYFGHNHNMAMIPALIGTFYAAAFVPMMAVPALEPKHKAISVYN 360
QY 361 LSAGMSNFLPAIVAVLLPWFSTIGVYIAYATYLLAFVLCAPFRVEQPGFSSAPYTEKA 420
DB 361 LSAGMSNFLPAIVAVLLPWFSTIGVYIAYATYLLAFVLCAPFRVEQPGFSSAPYTEKA 420
QY 421 LNIS 424
DB 421 LNIS 424

RESULT 2

Q82K49 PRELIMINARY; PRT; 408 AA.
ID Q82K49;
AC Q82K49;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative permease.
GN STM4434.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SCS1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McCalland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2";
RL Nature 413:852-856(2001).
DR EMBL: AE008908; AAL23254.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 408 AA; 44676 MW; DCE20DAAE075121B CRC64;

Query Match 28.3%; Score 625; DB 16; Length 408;
Best Local Similarity 33.3%; Pred. No. 5.3e-37;
Matches 134; Conservative 84; Mismatches 169; Indels 16; Caps 8;

QY 12 PLHLIMGYIAIVPMPTGDFELAFSLHYIKSLGTPPAEASAFLLYGLAALMSAWSGV 71
DB 16 PKFTWVSLCTFLFEMLGCIENSWSLAMLNPGDQAHIGOIFAGYIVAIVATITMSLGCVC 75
QY 72 AEITTPQKTMILGFVLCVHFVLFVGLGOANYGLILFYGIGLAVPLFLYFIVII 131
DB 76 VDVGPRKAVWTGRIYVLLASVAFNLALPSHDGALLVYMLRGVGPLVCYFIVRL 134
QY 132 HNVRSESSSALG--WYAVSVGIGVAGSYIPSTPIPMGEMGLMLALAFCPAGVIA 189
DB 135 -TOLDNHQOIGISLFWVYVNLGFTTIGPVASLIPELCHINVMWAGMKVALL--GVLF 192
QY 190 MISLRHVTKGHNHLPREK--PAELSRATVTLTYNRIIFLSSIVNIITLSLFGFAVI 247

DB 193 MLVIER-----NEFLPRTTPYVEKELSGAGISIMERPRLGLAVITINLGYGFAVYV 247
QY 248 MPMFVDELGFTTSEMLQVMAAFETTFISNIFMGIVAEKKGMVRIRWEGCLGMASSL 307
DB 248 LPLFLDLCK-HTTLEEMASIMGITFISNOVENIIFGWMGDKIGFRTQIIGSILITGVA TL 306
QY 308 AFYMPQYFGHNHNMAMIPALIGTFYAAAFVPMMAVPALEPKHKAISVYNLSAGKSN 367
DB 307 IVYVPMIMGHNYVAFMLAMCLMGAGLGFVPMPLVPMMAVPALEPKHKAISVYNLSAGKSN 366
QY 368 FLAPAIVAVLLPWFSTIGVYIAYATYLLAFVLCAPFRVEQPG 410
DB 367 FVGPAI--VSYLAGFTGVYMTMAGLITLFGSILVQFLKV--PG 406

RESULT 3

Q9HWG1 PRELIMINARY; PRT; 462 AA.
ID Q9HWG1;
AC Q9HWG1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Probable MFS transporter.
GN PA4233.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_Taxid=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Britman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AE004840; AAC07621.1; -;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane; Complete proteome.
SO SEQUENCE 462 AA; 48818 MW; C2178DCE2DB45F0 CRC64;

Query Match 8.2%; Score 181; DB 16; Length 462;
Best Local Similarity 22.6%; Pred. No. 2.7e-05;
Matches 100; Conservative 75; Mismatches 182; Indels 86; Caps 20;

QY 13 LHLIMGYIAIVPMPTGDFELAFSLHYIKSLGTPPAEASAFLLYGLAALMSAWSGV 71
DB 20 LSLVFAFRMLGEMFV-----LPVLTATYGGDLAGATPALIGLAIGAYGLTQAILOIPGTI 74
QY 72 AEITTPQKTMILGFVLCVHFVLFVGLGOANYGLILFYGIGLAVPLFLYFIVII 128
DB 75 SDRIGRPVIVG-----LLIFRAGALALANDSITMGVAGVLOGAGAGISAAVM 124
QY 129 VIHNV--RENSSSSALCWYAVSVGIGVAGSYIPSTPIPMGEMGLMLALAFCPAGVIA 178
DB 125 ALLSDLTREQHRTKAM-----AMIGMSIGV-----SPAAVAVLQPLVTLHFGHLGFLWFT 174
QY 179 LAFCPAGVYIAMIISLRHVTPGHNHLPREKPAELSRATVTLTYNRIIFLSSIVNIITLS 228
DB 175 AGMALVG--LLTLFVPOPDHMTQHRESSVARQALLPLTKHGDLLR----- 219
QY 229 LSSIVRIINTLSLFGFAVIMPMVDELGFTTSEMLQVMAAFETTFISNIFMGIVAEKM 288
DB 220 LDAGILILHAILMASF--VALPLALVHGGGLPKREOHMMVYLTALLVGFGAVPFIIVAEKK 278

QY 289 GMMVIRIRFGCLGMAASLAIYTPQYEGHHYMAAMI.PAIALGFEVAAPFMPAAVPALPE 348
Db 279 RRMARV----LYGAVATLLACELFEFLVFGHSLALVVGTV--VFETAFNLLERSIPSLV 331
QY 349 PK-----HGKAIVSYNLSAGSNFELPAIATVVLTPV-FSTIGVIAVYTLALYLAFF-LC 401
Db 332 SKVSPAGSGKGRAMGVYS---TSQFLGALAGGILIGMFMFHGHSNMFICCAVIALALMLA 387
QY 402 AFIRVEQGFSSA---PYTEKAL 421
Db 388 IAVTMREPPYVTSIRLLPAAL 410

RESULT 4			
082C53			
ID	082C53	PRELIMINARY;	PRT; 456 AA.
AC	082C53;		
DT	01-MAR-2002 (TREMBLrel, 20, Created)		
DT	01-MAR-2002 (TREMBLrel, 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel, 21, Last annotation update)		
DE	Putative transporter.		
CN	YP03169.		
OS	<i>Yersinia pestis</i> .		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	<i>Yersinia</i> .		
OX	NCBI_Taxid=632;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CO-92 / BIOVAR ORIENTALIS;		
RX	MEDLINE=21470413; PubMed=11586360;		
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,		
RA	Prentice M.B., Sebahlia M., James K.D., Churcher C., Mungall K.L.,		
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,		
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,		
RA	Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,		
RA	Leemond S., Moule S., Oyston P.C.F., Quail M., Rutherford B.,		
RA	Slimmons M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;		
RT	"Genome sequence of <i>Yersinia pestis</i> , the causative agent of plague,"		
RL	Nature 413:523-527(2001).		
DR	EMBL: AJ141155; CAC92404.1; "		
DR	InterPro: IPR003662; sub_transporter.		
DR	Pfam: PF00083; sugar_tr; 1.		
DR	Hypothetical protein; Complete proteome.		
QO	SEQUENCE 456 AA; 48790 MW; 74D65EF5495B4095 CRC64;		

	Query Match	8.0%;	Score 176;	DB 16;	length 456;
	Best Local Similarity	22.2%;	Pred. Mismatches 182;	Indels	Gaps
	Matches 99;	Conservative	86;	Misclatches 182;	Indels 78; Gaps 21.

Qy	1 MSRRNKKOMLGLPLHLIMWGY-IAIAVFMWGDDGCELAFLSHYIKSL-GTFPEASFAFTLYG	58
Db	1 MAMNNKMTPELRATRWGTGIVFSRLRMGMFNVLPVLITTYGNALSGASALLGIALGIYG	60
Qy	59 LAAALSANVSGVAELITPQKTMLIGEVLMCVFHWLEFLVGJGQANGYLILFYGIRGA	118
Db	61 LSQAIFQIPIFGLLSDRI-GRKPMIIIGGLL-VFALGSIIAALSDSIWGCIIL---GRAQL	113
Qy	119 YPLFLXSFYVIINHV-RSENSSSALGWMAVYSGICAGSYIPSFTTIPINGEKOTML	177
Db	114 GSGATAAVALMILLSDTREONRTKMAEFGVSFGTEPAFAA-----MVLGSPITYTHAF	164
Qy	178 ALAPCFAG-GVYAMISLRHVKT--PGHMHNLPREPKEAFELSRAVTLTYNRNIPLSSIYR	234
Db	165 GLQALEFWGAIITALLCGYITLTIVVSSANSRVINRESSMKWSKYVLAHNSRLKLNFGLIM	224
Qy	235 IINTLSLEGFAVIMPMFVDDELGFPTTSEMLOVMAAFEFTTIYS-----	277
Db	225 CLHIILMSSE-VALPOUMMAN-AGLAPA--QHVVVYLVLTMLVSPAAYVPFIYAEMKRMM	279
Qy	278 -NIIFGVIAEKKGMRVRVIRMECCCLGMAASSLAFYTMPPQIFGHNYMMAMPALALGTFVNA	336
Db	280 KQVFNGCYA-VLFIAEVLVMAF-----GQDLWI-IIAAGVOL--FFTA	317

[illegible]

```

RESULT 5
053459
ID Q53459 PRELIMINARY: PRT. 388 AA.
AC Q53459;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE NORa.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94379790; PubMed=8092836;
   Ng E.Y., Trucksis M., Hooper D.C.;
   "Quinolone resistance mediated by norA: physiologic characterization
   and relationship to flgB, a quinolone resistance locus on the
   Staphylococcus aureus chromosome.";
   Antimicrob. Agents Chemother. 38:1345-1355(1994).
RL Anticrib.
CC -1- SUBCELLULAR LOCATION: INFERIAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; S74031; AAB31949.1;
DR InterPro; IPR004734; Drug_reslst.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PFO0083; sugat_tr_1
DR TIGRFAMs; TIGR00880; 2_A_01_02.1.
SQ Transmembrane.
SEQUENCE 388 AA; 42323 MW; 6C63883817827EAF CRC64;

```

	Query Match	7.9%	Score 175;	DB 2;	Length 388;	
	Best Local Similarity	20.5%;	Pred. No. 6.2e-05;			
	Matches	85;	Conservative	77;	Mismatches 190;	Indels 62; Gaps 15;
Oy	5	NKQWLGRLHLIMGYTAIAVPMTGGDFELAFLSHYKSLGFTPAEASFAETLLYGLAALS	64			
Db	2	NKQIFVL-----YFNIFLFLIGLIVLPVPLVLLDKLIGSDGLGLVAFAALSQMIT	54			
Oy	65	AMVSGVAEITPQKTMILGFLMKCVPHVLLVFGIGQANVGLLILFLYICRLATPRLFLX	124			
Db	55	SFPGGTGLADKGLKLLICGLILFLVSSEPMFV-----GNHVSFLMLSRVIGMSGGMVP	110			
Oy	125	SEIVIIINHVRSENSSALGWYMAVSVGIGVAGSYTPEFTPIRMGEMTLLALAFCA	184			
Db	111	GYTGLIADISPHQAKKNQGYMSALIIINSF-ILGCGIGFMAEVSHRM-----PFLYA	162			
Oy	185	G--GVIAIMS--LRH--VKTPGHMHNLTPREKFAELSRAVTLLTYNBNIFLSSIV-RI	235			
Db	163	GALGILAFIMSYIVLIHDPKSTTSQFOKLEPQ-----LTKIKMKVETITVLIITL	212			
Oy	236	INTLSLFGRAVINPMPEVDELGETTSE-WLQYMAAFPTTIPSNNFMCIVAAKMGMMRI	294			
Db	213	VLSPGLSAETLYSLYTDKVMYSPKDISIALTGGIGFALPOLYIFD-----KFM	263			
Oy	295	RMFGCLIGMAASSLAF--YUMPQYEGHNYMMYMLPAIALGTFVAAFVPMAAV--FPALD	348			
Db	264	KYFSLLTFLIAMSLLXSVVVLILLVRAANDWMSIML--ISLVEIGDMIRPALTNFESNIA	321			
Oy	349	PKHKGAISVYNLSAGMSNFLAPATAVULL-----PMESTGIVAIYATATYLL	396			
Db	332	GERQGFAGLGINSFTSMGNEIFGRPLLAGLFDVHIEAPRIYMAIGVSLAGVAVILI	375			

RESULT 6
003325 ID 003325 PRELIMINARY; PRT; 388 AA.
AC 003325;
DT 01-NOV-1996 (TREMblrel, 01, Created)
DT 01-NOV-1996 (TREMblrel, 01, Last sequence update)
DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE FLUOROQUINOLONE resistance NORAI199 protein (ALLELE NORAI199) (Allele
NORAI199(WT)).
GN NORAI199.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
CC Staphylococcus.
CC NCBI_TaxID=1280;
RN [1]
RP ALLELE NORAI 1199, SEQUENCE FROM N.A.
RC STRAIN-SA-1199B, AND SA-1199;
RX MEDLINE=93297926; PubMed=8517696;
RA Kaatz G.W., Seo S.M., Ruble C.A.;
RT "efflux-mediated fluoroquinolone resistance in Staphylococcus
aureus."
RL Antimicrob. Agents Chemother. 37:1086-1094(1993).
CC -1- FUNCTION: INVOLVED IN FLUOROQUINOLONE RESISTANCE. MAY CONSTITUTE A
MEMBRANE-ASSOCIATED ACTIVE EFFLUX PUMP OF FLUROQUINOLONES AND
OTHER DRUGS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- INDUCTION: FUNCTION OF NORAI199 PROTEIN AS MULTIDRUG EFFLUX
TRANSPORTER IS INHIBITED BY RESERPIN.
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC DRUG RESISTANCE TRANSLOCASE
FAMILY.
CC -1- SIMILARITY: TO MULTIDRUG RESISTANCE PROTEIN FROM BACILLUS SUBTILIS
(P33449).
CC -1- CAUTION: THE SEQUENCE SHOWN IS THE SAME FOR ALLELE NORAI199 AND
ALLELE NORAI199(WILD TYPE). STRAIN SA-1199 CONTAINING ALLELE
NORAI199(WT) IS FLUROQUINOLONE-SUSCEPTIBLE. INCREASED
TRANSCRIPTION, AND NOT GENE AMPLIFICATION, OF NORAI199 IS THE
BASIS FOR FLUROQUINOLONE RESISTANCE IN STRAIN SA-1199B.
DR EMBL; M80252; AAA16158.1; -;
DR EMBL; M97169; AAA26658.1; -;
DR InterPro; IPR004734; Drug-resist.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR00880; 2.A.01.02; 1.
KW Transmembrane; transport; Antibiotic resistance.
SQ SEQUENCE 388 AA; 42236 MW; 81351490F77BD663 CRC64;
Query Match 7.9%; Score 174; DB 2; Length 388;
Best Local Similarity 21.0%; Pred. No. 7.3e-05;
Matches 87; Conservative 75; Mismatches 189; Indels 64; Gaps 16;
QY 5 NKOWGLPLHLIMGYIANVFMGDDGELAFSLHYIKSLGFTPAEDASPAFTLYGLAALS 64
DB 2 NKOLIVY-----YFNIFLFLGIGVIVPLPYLKDILGTLGDLGLVNAAFALSQMII 54
QY 65 AMVSGVAEITTPKOTMLIGFVLCVPHVLELVFGLOAGNGLLLPYGNGLAYPEFLY 124
DB 55 SPEGGTADKLGKLLICIGIL--FVSSEFMFAIQ-NELIIMLSRVIGGMSAGMWP 110
QY 125 SFVIVIIHNVRSESSSALGYMAVSVGIGVAGSYIPSEFTPIPMGEMGLMLAFCFA 184
DB 111 GVTLGIADISPSHQKAKNFYMSAIIINSGF-ILGPGIGGFAEYSHM-----PYFA 162
QY 185 G--GVIMT-----SLRHVKTPGHHNHLTPREKFAELSAVTLVLTNRNIFLSSIV-R 234
DB 163 GALTILAFINSIVLIHPKRVSTNG-FQKLEPQ-----LLTKINMKVFITPVILT 211
QY 235 IINTLSLFGFVAVIMPMFVDELGFTSE-MQVNAFAFFTTFININMGIAEKMGMKV 293
DB 212 LVLSFGISAEETLSLTADKAVNSPKDISAIIIGGIGFALGFQIYFFD-----KF 262
QY 294 IRMEGLGMAASSLAF--YYMPOYFGHNVMAMIPALIGTFFVAALVPMVA--FPAL 347
DB 263 MKYFSELTFIAMSLLYSVIYLVLIADGYWTIMV--ISFVFIQFDMKIRPAIINYSNI 320

QY 348 EPKHGAIAISVYNLSAGNSNFLAPAIIVLL-----PWSTTIGVIATYALYLL 396
DB 321 AGDRGFAGCINSTFTSMGNFISGLIAGALFDVHIAPYIAVYAGVAVIYL 375
RESULT 7
034502 ID 034502 PRELIMINARY; PRT; 444 AA.
AC 034502;
DT 01-JAN-1998 (TREMblrel, 05, Created)
DT 01-JAN-1998 (TREMblrel, 05, Last sequence update)
DT 01-MAR-2002 (TREMblrel, 20, Last annotation update)
DE YVKA.
GN YVKA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
CC Bacillaceae; Bacillus.
CC NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Soldo B., Rivoita C., Reynolds S., Manuel C.,
RA Karamata D.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berrero M.C., Bessieres P., Bolotin A., Borchert S.,
RA Broulier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Broutlier S., Bruschi C.V., Caldwell B., Capuano N.J., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dufterloft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Funa S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konlingstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Prescecan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Roche A., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Takemura S., Vandenbol M., Vannier F., Vassarotti A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Winters P., Wipert A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AF017113; AAC67266.1; -;
DR EMBL; Z99122; CAB15538.1; -;
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane; Complete proteome.
SQ SEQUENCE 444 AA; 47797 MW; 6A27F4EFB836CE8E CRC64;

FT	TRANSMEM	365	385	POTENTIAL.
SQ	SEQUENCE	386 AA:	42068 MW:	2892868E56D1BFC1 CRC64:
	Query Match		7.4%:	Score 164.5; DB 17; Length 386;
	Best Local Similarity		20.0%:	Pred. No. 0.00035;
	Matches	82; Conservative	91; Mismatches	189; Indels 47; Gaps 15;
OY	15	LIMGVIATAVFNTGGGFEIATLSHIKSLGTFPRAEASFATLTXGLAALASAMVSGVAET	74	
Db	10	YIWM---ITFTFTMLGGVFIAPIMAIYAQTGAANLEIGLFGSFALARTVAQIPVGLSDI	67	
OY	75	ITPQKTMILGFIYLMGVFHVLF--FLVEGGLQAVNGLILFYGSRGLAVPLFLTFSFVIID	132	
Db	68	YGRKFYIYCGTFYGVSTILMNFVSTVLG---FLVIRITGTFSAFVPVPAGSYIAATLP	124	
OY	133	NVRSNSSSALGQWYAAVYSGVAGSYTPSTPIPMGEMLTLMALAECPAGVYAMIS	192	
Db	125	KTR-----LGEYMGIDNSAI--TLGFGIGPFGIGILDMWYGIKMPFYCGELITAAI-	175	
OY	193	LKHVKTGPHMHLTPREKPAELSRAVTL---LYNNRNIFLSSYIRIITL---SLFGTA	245	
Db	176	ISYMKLEDIVFNKN-KETI-DYKAKISTLTFSEFELKRNNSSSFLINVSVMNAGIAYTL	233	
OY	246	VIMPMF---VDELGETTSEMLQVAAAEFTTIFSNIF--WGIVAEMKGMARVIRWFCG	299	
Db	234	ALYATINYNTTISQVGFMA-----LTNIMALLQRSRGLKLDKGNIMIT-----	278	
OY	300	LGMAASSLAFYIMPOYFGHNWYMMAMIPALACTGYAAEVPMAAVPALEPKHGAISYV	359	
Db	279	IGIFITISFGMYLLSTSTFELLTALSLITIAVSSSISSTFATSLAVKDIPTHRKGEAMGLF	338	
OY	360	NLSAGMSNPFLAPAVLVLLPMWSTIGGVYIAVATLVLAFLVCAFIREVQ	408	
Db	339	TTSINIGMFIC-AVSFGFLADILGLIANNMKFSATSIYVGLISYLRTER	386	

Db	1	MIAKFPFMY-----ETTRPQKNALFSAMLGVPDGFDFPMILFYIMYLKADGLGDTDEGA	56
Oy	101	-----GOANNGILLPFPGINGCLAPLFLYCFYVIYTHNVSSSSALGWTMA	148
Db	57	FLATAAFIGRPFPGALFGLADKFGKR-----PLMMWSIVAVSVGTGSLGASGYIMTLIS	112
Oy	149	VYSVGIGVAGSY-----IPSFPIPMOEGTLM	176
Db	113	RFIVMGHAGGEYACASITYAVESWPKHLKSKASAFVSGFGIGNITAAIYMPSPFE-AYGM	171
Oy	177	LALAFCEAGVYIAMISLR-----HVKTPGHMHNLTPREKFAELSRAYTLLYTIN	224
Db	172	RAAFVGLGLPVLLIYIYIRAPRESKEWBEAKLSGP-KHS-----QSAVSFSLSKMGLF-N	226
Oy	225	RNIPLSSIVRIINLSLFG-----FAVIMPMFVDELGFTTSEMLOQVAAAFPTTISENIF	260
Db	227	RAQFPPLTCEVIFLSEFGANWPIJFGLLPYLTGE-GEFTDGVVSNLTAAAFGVIGLNI	285
Oy	281	WGIVAEKMGWRIVRWFGCLGMASSSLAFYMPQYFGHNWM--AMIPALALGTVEAAF	337
Db	286	WGLCADRIGLKTTP-----SIGLMSPLFIPLPRIPDNTLILGLACFLGMLATVAGVGL	341
Oy	338	VP--MAAVPALBPKHKHGAISVYNLSA-----GMSNPLAPAIAYVLLPMS	382
Db	342	VPKELYDFP-LEVRIGLTGL-IYNLAATSGTENSMAATWLGITMGALATFIYAPWTA	399
Oy	383	TIGVY	388
Db	400	TILLII	405

```

RESULT 10
082032
ID 082032 PRELIMINARY; PRT; 426 AA.
AC 082032;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative sugar transport protein.
GN STM1132.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_Taxid=602;
[1]
SEQUENCE FROM N.A.
RP STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Potwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
"Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
EMBL: AE008749; AAL20063.1; -.
DR InterPro: IPR003662; sub_transporter.
PFam: PF00083; sugar_tr.1.
KW Sugar transport; Hypothetical protein; Complete proteome.
CO SEQUENCE 426 AA; 46492 MW; F5E7C8D65FB33857 CRC64;

```

RESULT	ID	Q827M8	PRELIMINARY;	PRT;	426 AA.
AC	Q827M8				
DT	01-MAR-2002	(TREMBLrel. 20, Created)			
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Putative transporter.				
GN	STY1169.				
OS	Salmonella typhi.				
CC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OX	Salmonella.				
NCBI_TaxID=601;					
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-CT18;				
RX	MEDLINE:21534947; PubMed:11677608;				
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,				
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,				
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,				
RA	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whithead S., Barrrell B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18."				
RL	Nature 413:848-852(2001).				
DR	EMBL; AL627269; CAD08257.1;				
DR	InterPro; IPR000719; Euk_pkinase.				
DR	InterPro; IPR003662; sub_transporter.				
DR	Pfam; PF000083; sugar_tr.1.				
DR	PROSITE; PS00107; PROTEIN_KINASE_APP; UNKNOWN.1.				
DR	Hypothetical protein; Complete proteome.				
QO	SEQUENCE 426 AA; 46546 MW; 6B685103C92f79FE CRC64;				

Query Match	7.3%	Score 162.5;	DB 16;	Length 426;
Best Local	Similarly	Pred. No. 0.00053;		
Matches	91; Conservative	68; Mismatches 150;	Indels 117;	Gaps 19

Query Match	7.3%	Score 162.5	DB 16	Length 426
Best Local Similarity	20.6%	1 Pred. No. 0.00053		
Matches 89; Conservative	66;	Mismatches 147;	Indels 131;	Gaps 17.

```
QY      59 LAALSAWSGVVALEITTPQKMLI-----GFLMCFVHLYLV--EGL----- 1000
```

```

QY      59 LAALSAWNSGVVAELITPQMTLI-----GFLWCVFHYFLV---FGL----- 100
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 1 MIAKEFPWYS---ETTRPQKNALFSAMLGVDGDFMILFYIMYLKADIGLADMEGA 56
 Oy 101 -----GOANGYLILFYGIRGLAYPLFYSPFIWIIHNVRSSSSALGMWYA 148
 Db 57 FLATAAFITGRPGALFGLADKFGK-----PLMMMSIYAVSVGIGLSGLASGVIMLTIS 112
 Oy 149 VYSGIGVAGSY-----IPSFPIPIGEMGTLM 176
 Db 113 RFTYGMAGKACASTYAVESMPKHLKSKASAFVSGFGIGNIITAAVMPSPAE-AYGM 171
 Oy 177 LALACFAGGAVIAMISLR-----HKYTGCHMNLTPREFALISRA 217
 Db 172 RAAPFVGLPLVLLIYIRARAPESKWEBAKLSGLGHSQASMSVSLSMKGF----- 225
 Oy 218 VLLTYNRIPLFSIYRIINTLSFG-----PAVIMPMVDELGFTSMLQVMAAFPT 273
 Db 226 -----NRAPFLTLCVFLVFSIFGANNPIFGLLPTIAGE-GEPTGVVSNLMTRAAG 278
 Oy 274 TFSNIFMGIVAEKMGWMMVIRMFGLGMAASSLAFLYMPQYFGHNYM---AMIPAL 330
 Db 279 TVLGINVWGLCADRIGLKRTF-----SIGLMSFLIFELFRIPQDNYLLGLACFLGMAT 334
 Oy 331 GFVNAFV-MAAVPRALEPKHKAISVYNSA-----GMSNFLAIAIV 375
 Db 335 NVGVGGLVRFYDFP-LEVRLGLTGL-ITNLAATSGTFNSMAATWLTITMGLAALT 392
 Oy 376 VLLPMFSTIGVVI 388
 Db 393 IYAFWTATILLII 405

RESULT 12

P94376 PRELIMINARY: PRT: 400 AA.

AC P94376: 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein yx1h.
 GN yx1h.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RP SEQUENCE FROM N.A.
 RC STRAIN-BGSC 1A1;
 RA MEDLINE-97124196; PubMed=8969509;
 RA Yoshida K., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT *Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the 11c and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacX region.*;
 RL Microbiology 142:3113-3123(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA MEDLINE-98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borcher S.,
 RA Boriss R., Boursier L., Brans A., Braum M., Brignell S.C., Bron S.,
 RA Braillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Chol S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Fougere D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karmali D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesen E., Puig P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate I.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT *The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis*;
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: D83026; BAA11739.1;
 DR EMBL: 299123; CAB15890.1;
 DR InterPro: IPR003662; sub-transporter.
 DR Pfam: PF00083; sugar_tr.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 400 AA: 43963 MW: 185A568E76E4D38 CRC64;

Query Match 7.3%; Score 161.5; DB 16; Length 400;
 Best Local Similarity 21.5%; Pred. No. 0.00059;
 Matches 93; Conservative 70; Mismatches 158; Indels 111; Gaps 18;

Oy 18 GYIAIVFMTGDFELAFISHYIKSGFPPAEASFATLYGLAALSAVWGVAEII-T 76
 Db 33 GYIYVVF-----SLYEDLHFSYGAIGITLGSYQVILRPPLGLSDILS 81
 Oy 77 POKTMLI-GFVLMCVFHVFLVFGLGQAVYGLT--LFGIRGLAYPL-FLYSPFIYV 130
 Db 82 LRQQLIAGFGSVLSLSLFLMD--SPFVLARLRFAGITASVWATILYA----- 132
 Oy 131 IHNVRSESSSALG--WYNAVYS-----VGIVAGSYIDSFITPIGEMGTMLALAF 183
 Db 133 -HYFNNGNSKAMGIMQFETVMPQFASIVFCGLAAH-----LGRVPEWMAALASA 183
 Oy 184 AGGVIMISLRHKTPGHHNLTPREKPAELSAVTL-LYTNRIPLFSIVRIINTLSF 242
 Db 184 AGLVICCFIODPSAPPANNGTIRVNOYIKDTLPKLTFT---ILSTPAHVAFLTYV 239
 Oy 243 GFVIMPMFVDELGFTSEMILQVMAAFPTTTFNSIFMGIVAEKGMWVIRMFGLGM 302
 Db 240 GFV-----PLYMNLGMDDELWMSAF-----LPH 267
 Oy 303 AASSLAFLYMPQYFGHNYMAMIPALALGTVAAYPMAA-----VP 344
 Db 268 AATLSFVFL-RFTSRIVASVMLISFAVGVCLLYVFLAALTFTVCTHACIGLAFVF 326
 Oy 345 PAL-----EKHKGAISVYNLSAGMSNFLAPAVLVLPWFSTIGVIAVATLYLL 396
 Db 327 PLLSHVVEISSARLMSWGFYOSFYALGIFLPLAGKIAOLIGLAGVFGAGSLARA 386
 Oy 397 AFVLCAFIRVEQ 408
 Db 387 AF-----FVMAAQ 394

RESULT 13

O82PR3 PRELIMINARY: PRT: 411 AA.

AC O82PR3: 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative MFS family transport protein.

GN YDIN OR STM1360.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SSSC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2".
RL Nature 413:852-856(2001).
DR EMBL: AE008759; AAL20285.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 411 AA; 44788 MW; 5E544C218B01B884 CRC64;

Query Match 7.3%; Score 161.5; DB 16; Length 411;
Best Local Similarity 22.0%; Pred. No. 0.0006; Indels 93; Gaps 21;
Matches 97; Conservative 71; Mismatches 179;

QY 1 MSRRNKKQMLGPLHLINGYIAIAVFMTC--DGFELAFSLHYI-KSLGTPPAEASFATLY 56
DB 1 MSQKKAR--NMPYLAVICIFYSYFL--HGMVSITTLAQNMTSLAQKSTDSAGIAYLISG 56
QY 57 YGLAALSAWVSGVAEITTPQKTMILGFVLMCVFHVLFVFGIGQANYGLILFYGIRG 116
DB 57 IGLGRVLSLFFGVLSDFKFGRRRIITLLGAVLYMLF-----FEGIP- 96
QY 117 LAAPLFYSFTVIYIHVRSSENSSALGWYAVYVSGVAGSYIPSTPIPMGEMGT-- 174
DB 97 -ASNNIAITLAVCVV--ANSALDTGCTPALMECPKASGSAV--TLKAMVSFCOMI 151
QY 175 -----LMLALFCFAGGVIAIMISLRHYKT--PGHMNLTPREKFAELSRVTL 220
DB 152 YPLVLSALLVNHMYGAAYVIRPILFVLITLMLKSRPQLVD-----ASIAKELPQ 204
QY 221 LYTRNRTFLSSIVRIITLSLFG-FAYI---MPMFYDELGFTTSEMLQWMAAFETTF 276
DB 205 MNSPPLWMLGVAISVLEGVAAFTFVYIVVMPKYMAAFGMAESDALKTIISYSGSLV 264
QY 277 SNIFMGIVAEKMG--WMRYIRMEFCIGMAA--SSLAFYMPQYFGHNYMMAMIPAI 331
DB 265 CVFIFALLKKMVRPIANFN---AGLATITAAIILYLP-----PLICNAG 309
QY 332 TTYAAAEVPM-----AAVFPALPEPKHGAISVYNLSAGMSNFLAPAIIVLLPWFSTI 384
DB 310 AFVIGFSAAGCIIQLGVSVMSEFPKSKAKVTISYMMGVANFLPLIT-----GYLSTI 365
QY 385 GVVIATYATLLAFVLCAT 404
DB 366 G--LQYITLLDFAFALLFTI 383

RESULT 14

Q82PG6 PRELIMINARY; PRT; 428 AA.
AC Q82PG6;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative transport protein.
GN STM1543.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SSSC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2".
RL Nature 413:852-856(2001).
DR EMBL: AE008759; AAL20285.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 428 AA; 46461 MW; 20901CC429724AA3 CRC64;

Query Match 7.2%; Score 159.5; DB 16; Length 428;
Best Local Similarity 20.3%; Pred. No. 0.00087; Indels 143; Gaps 19;
Matches 99; Conservative 73; Mismatches 173;

QY 1 MSRRNKKQMLGPLHLINGYIAIAVFMTC--DGFELAFSLHYI-KSLGTPPAEASFATLY 57
DB 1 MKRTMGWLVY-----FLFTIYMLNMDRSLSTTAPLKEKELGFNAEMGMTFSAF 53
QY 58 GLAALSAWVSGVAEITTPQKTMILGFVLMCVFHVLFVFGIGQANYGLILFYGIRG 117
DB 54 FIVGLAFNFGVMAKSDKVPKRVFLAALLMSV-----FCGLTGL 93
QY 118 APLFLXSFIVY-----TIHN-VASENSSALGWYAVYVSGVAGSYI 161
DB 94 VTGLMTLVIRVLEGMAEGPVSAAGKTIINWISRKESATAIGFSGSPGGAVSG-- 150
QY 162 PSFTPIPMGEGFTLMLALFCFAGGVIAIMISLRHYKT--PGHMNLTPREKFAEL 213
DB 151 -----PL--VGLLALSGLRRPAGIITFLFGVAVVLLMYTIVSKPRMSKRLAPEEID- 201
QY 214 LSRAVTLTYTRNRTFLSSIVRIITLSLFGFVAVIMPMEYDELGFTTSEMLQWMAAFET 273
DB 202 -----FENHEDVITLSDDRATPSTL--GYMKQPMVMTTLAFTFSYNYI----LEFFL 247
QY 274 TIPS---NIFMGIVAEKMGMRVIRW--FGCLGMAASSLAFYMPQYFGHNYMMAMIPAI 329
DB 248 TWEPSTYNHSLHDIDIKETISATYIAPVYIGAIAGVAGVCSDVYIRITGN--ALLSRLLI 304
QY 330 LGTFVAAFPMAAVFPALPEPKHGAISVYNLS-----AFIR 362
DB 305 LGVGLAAGAAVAVSGVSR--IGSATITMSVSLFLILYLTGPYIMAVIQDVHKDKVSV 362
QY 363 -----AGMSNFLAPAIIVALLPWFSTIGVVIATYATLLAFVLC-----AFIR 405
DB 363 GGMHGLANISGIILPVT-----GFIQVFSKDYAFAYLAGAIAIVSLLVFFV 412
QY 406 VEQPGFSS 413
DB 413 VKSKGFK 420

RESULT 15

Q82716 PRELIMINARY; PRT; 428 AA.
AC Q82716;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Membrane transport protein.
GN STY1519.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=601;

RP SEQUENCE FROM N.A.
RC STRAIN-CT18;
RX MEDLINE-21534947; PubMed-11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Dougan G., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogan A., Larsen T.S., Leather S., Mouton S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant *Salmone*
RT *enterica* serovar Typh CT18.";
RL Nature 413:848-852(2001).
DR EMBL: AL627270; CAD01773.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Complete proteome.
SQ SEQUENCE 428 AA; 46519 MW; 5DBFE2EBD6E4AF74 CRC64;

Query Match 7.2%; Score 159.5; DB 16; Length 428;
Best Local Similarity 20.3%; Pred. No. 0.00087;
Matches 99; Conservative 73; Mismatches 173; Indels 143; Gaps 19;

QY 1 MSRRNNKQWLGRLHLIMGYIAIVFMTG--DGEELAFLSHYI-KSLGTPPAEASFAETLY 57
DB 1 MKRRTMGWLV-----FLFLIVMLNMYDRSALSTIAPLIEKELGFNAEKGMIFFSAF 53
QY 58 GLAALSAWVGVVAEITFPQKTMIGFVLCVFEHVLFLVFGLGQANYGLILFYGIRGL 117
DB 54 FICYALFNFIDGWSADKVGKTYFLAALLMSV-----FCGLTGL 93
QY 118 AYPLFLXSLIV-----IIHN-VRSNSSSALGWMYVYVSVGIGVASYI 161
DB 94 VTGLWMTLVRYLFGNAEGPVSAGNKKIINMWSRKSATAIGIFSAGSLPGAVSG--- 150
QY 162 PSFTIPIGEMGFLMLALFCFAGVYAMISLRHV-----KTPGHMNLTPREKFAE 213
DB 151 -----PI---VGLLALSLGWRPAFGIIFLGLVWVLLMYFIVSDKPTMSKRLAPERID- 201
QY 214 LSRATVLLTYNNRIFLSSIVRIINLTLSPFAVIMPMFVDELGFTTSEMLOVMAAFPT 273
DB 202 -----FENHEDVILSDGGRATPSL--GYVMKQPMVWATTLAFPSNYI---LFFFL 247
QY 274 TIFS---NIFMGIVAEKMGMMRYIRN-FGLGMAASSLAFYTMPQYEGHNYMMAMIPATA 329
DB 248 TWPPSTLNSLHDIKEISIAIVIPWIGAIMVGLGVCSDVIYRITGN--ALLSRRLI 304
QY 330 LGTFVAAEFVMAAVFPALPKHGAISVYNS----- 362
DB 305 LGVCLAGAAVAVAGSTVST--IGSAITLMSVSLPLLYITGPYMAVIOVDVHKDKVGSV 362
QY 363 -----AGMSNFLAPALAVLLPWFSTIGVIAVATAYLALFVLC-----AFIR 405
DB 363 GGAMHGLANISGIIIGPLVY-----GFIYQFSGKYDAFYLAGAIAIVSSLLVTFV 412
QY 406 VEQPGFSS 413
DB 413 VKSKGFKA 420

Search completed: March 13, 2003, 16:55:16
Job time : 31.0265 secs

ettu